Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

The Influence of Irrelevant Information Upon Complex Visual Discrimination: M. H. HODGE 1
Effects of Instructions on the Transfer of a Conditioned Response: H. H. LINDLEY 6
Additivity of Cues and Transfer in Discrimination of Consonant Clusters: F. RESTLE 9
Discrimination in the Squirrel Monkey as a Function of Deprivation and Problem Difficulty: R. C. MILES 15
Effect of Preliminary Trials on Rate of Conditioning in a Simple Prediction Situation: D. LABERGE 20
Relation of Stress and Differential Position Habits to Performance in Motor Learning: A. CASTANEDA AND L. P. LIPSITT 25
The Differential Effects of Word and Object Stimuli on the Learning of Paired Associates: C. C. WIMER AND W. E. LAMBERT 31
Decision-Making Behavior in a Two-Choice Uncertain Outcome Situation: S. SIEGEL AND D. A. GOLDSTEIN 37
Extinction After Acquisition Under Different Verbal Reinforcement Combinations: A. M. BUCHWALD 43
Performance Decrement at an Audio-Visual Checking Task: W. E. KAPPAUF AND W. E. POWE 49
Performance in Different Segments of an Instrumental Response Chain as a Function of Reinforcement Schedule: K. P. GOODRICH 57
Replication Report: Meaning Established by Classical Conditioning: A. W. STAATS, C. K. STAATS, W. G. HEARD, AND L. P. NIMS 64

American Psychological Association

Vol. 57, No. 1 January 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor; HELEN ORR, Promotion Manager; SADIE J. DOYLE, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 57, NO. 1 JANUARY, 1959

THE INFLUENCE OF IRRELEVANT INFORMATION UPON COMPLEX VISUAL DISCRIMINATION [1]

MILTON H. HODGE [2]

University of Virginia

In a recent study Archer (1954) systematically manipulated the amount of relevant and irrelevant information present in the visual identification of geometric patterns. The *S*s were required to respond to certain dimensions arbitrarily designated as relevant. Other dimensions for which no response was possible were considered to be irrelevant. Under his conditions the amount of relevant information was the only variable found to be statistically significant. This raises the question as to the conditions under which the amount of irrelevant information is an effective variable. One such possibility is where the irrelevant information acts to increase the probability of competing responses. In the Archer experiment such competition was highly unlikely, since the irrelevant information was neither relevant to discrimination of any of the stimulus patterns, nor likely to be confused with the relevant information; it simply acted as "noise." However, in a situation in which irrelevant information may be relevant at another time, i.e., serves as a discriminative stimulus for another response under different conditions, competition is likely to be enhanced.

In an experiment (Hodge, 1957) preliminary to the one to be reported, such a condition was investigated. For one group the same stimulus dimensions were relevant cues in certain stimulus patterns, but irrelevant in others. This irrelevancy condition, called the sometimes-relevant group, was compared with replications of two (0-bit and 2-bit) of Archer's irrelevancy conditions.

The results of this preliminary experiment showed that differences among the irrelevancy conditions were not statistically significant, but in general, longer latencies and a greater number of errors were associated with the sometimes-relevant group. A significant interaction found between irrelevancy and practice suggests that the differences among the irrelevancy groups tended to be greater at the beginning of practice

[1] This article is based on a dissertation submitted to the Graduate Faculty of the University of Virginia in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to L. Starling Reid for his guidance and encouragement throughout the investigation. The study was supported by contract No. DA-49-007-MD-537 between the University of Virginia and the Office of the Surgeon General, Department of the Army.

[2] Now at Mary Washington College of the University of Virginia.

than at the end. Although by no means conclusive, these findings suggest that irrelevant information which is sometimes relevant has a detrimental effect upon complex discrimination performance.

The most plausible reason as to why larger differences were not found is that an insufficient amount of irrelevant information was employed. It was the purpose of the present experiment to determine whether additional irrelevant information would produce greater performance decrements in the sometimes-relevant condition. In addition, two further questions were asked: (*a*) whether the effect of the irrelevant information is increased as discrimination of the patterns is made more difficult, and (*b*) whether the effect of irrelevant information is reduced by practice in the discrimination task.

METHOD

Subjects.—The *S*s were 30 male students enrolled at the University of Virginia.

Apparatus.—The apparatus consisted of a set of response switches, a film projector, and recording devices. The *S* sat at a desk shielded from *E* by a large wooden partition. Across the width of the desk top, eight bidirectional, lever-action switches were mounted in groups of two. Thus, with eight switches and two directions for each switch, there were 16 response alternatives. The various alternatives were labeled with symbols representing the stimulus dimensions described below.

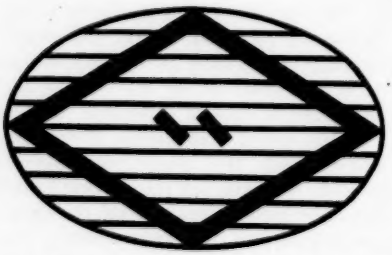

FIG. 1. Sample stimulus pattern (Form A).

The stimuli to which *S*s responded were projected on a white, translucent screen, 11¼ × 9¼ in., mounted in a window cut in the middle of the wooden partition. The stimulus films were projected automatically with a Dunning Animatic projector, adjusted to alternate between a ½-sec. and a 5½-sec. exposure time. A Standard Electric stopclock, calibrated in units of .001 min., was used to measure the latency of *S*s' responses. The clock was activated at the onset of the stimulus and was stopped when *S* moved one of the response switches. A panel of lights indicated to *E* which response *S* had made.

Stimulus material.—The stimuli, photographed on strips of 16-mm. color film, were complex geometrical figures which could be varied along 10 dimensions (Fig. 1). Each dimension had two values. In terms of the rules under which *S*s were instructed to respond, 16 combinations of the dimensions were possible. The dimensions and their values are as follows: (*a*) *Form.* The figures were elliptical in shape; one of the forms approximated a circle while the other resembled a very flat ellipse. These will be referred to as Form A and Form B, respectively. (*b*) *Size.* Half of the figures were exactly twice as large in size as the other half. (*c*) *Kind and number of marking.* Some of the figures contained crosses in the middle of the form and the remaining figures contained dashes. There were either one or two of the same markings in a given figure. (*d*) *Kind and brightness of border.* The stimulus figures could be enclosed in either a diamond-shaped or rectangle-shaped border. The borders were either filled in with black or they were left unfilled. (*e*) *Kind and stroke-width of hatching.* Vertical hatching was superimposed on some of the figures and horizontal hatching on others. The stroke-width of the hatching was either wide or narrow. (*f*) *Color and brightness of form.* Half of the figures were red and half were blue. These colors were either light or dark.

Besides the differences among the stimulus figures listed above, the shapes of Forms A and B were varied systematically in order to make difficulty of discrimination a variable. This was accomplished by holding the major axes of the forms constant while varying their minor axes in discrete steps. By this means, four different values of both Form A and Form B were obtained, a total of eight distinct forms.

Experimental design.—In all experimental conditions *S*s had to identify the stimulus patterns on the basis of the various stimulus dimensions. Two of the dimensions, form

and size, were present in all patterns. They will be referred to as the *primary* dimensions. The *secondary* dimensions, kind and number of markings, kind and brightness of border, kind and stroke-width of hatching, and color and brightness of form, did not occur in all patterns as relevant dimensions (see Fig. 1). Marking stimuli were invariably paired with the large Form A, border with the small Form A, hatching with the large Form B, and color with the small Form B. These combinations constituted the relevant stimuli of the experiment. In identifying the figures *S*s were required to judge which values of the primary dimensions were present and, on the basis of this judgment, to decide which pair of secondary dimensions was relevant. Since each secondary dimension had two values, there was a possibility of four responses after *S* had identified the relevant pair. The identification was completed when *S* moved the response switch corresponding to the values present in the stimulus pattern. Each of the 16 positions of the 8 switches was labeled with a symbol representing one of the 16 patterns.

The major variables of the experiment were: number of irrelevant dimensions in the stimulus figures, difficulty of form discrimination, and amount of practice on the task. Three groups, each with a different number of irrelevant dimensions, were run. In the simplest condition ($N = 10$) one pair of irrelevant secondary dimensions was added to the relevant dimensions. For example, hatching was added to the large Form A figures and marking added to the large Form B figures. The color dimensions were added to the small A figures and border to the small B figures. By use of this counterbalancing, *S*s were prevented from identifying the figures solely on the basis of the secondary relevant dimensions, i.e., they were forced to make use of the primary dimensions. In the second irrelevancy condition ($N = 10$) two pairs of irrelevant secondary dimensions were added to the basic stimulus figures. Again the irrelevant dimensions were added in such a way as to ensure *S*s' use of both the primary and secondary dimensions. The figures of the third irrelevancy condition ($N = 10$) contained all of the secondary dimensions (three pairs of irrelevant dimensions).

The second major variable of the experiment was the difficulty of the form discrimination. The four values of both Form A and Form B provided four levels of discrimination difficulty. The values were introduced randomly into all stimulus figures. Thus, each *S* performed under all levels of form discrimination.

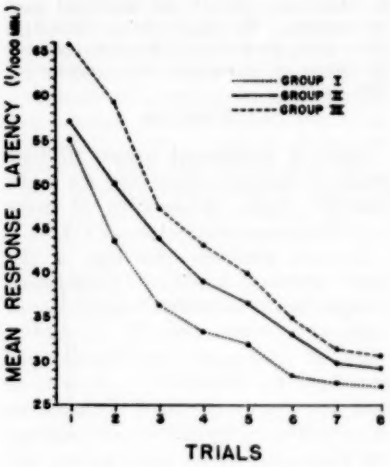

FIG. 2. Mean response latencies (in .001 min.) as a function of number of irrelevant dimensions and stage of practice.

A third major variable was practice. After completion of preliminary training, each *S* was given eight trials with an experimental sequence appropriate to his irrelevancy condition.

Procedure.—During preliminary training all *S*s were read instructions about the stimulus material and the task. The stimulus dimensions were illustrated by various sample figures projected on the screen. The *S*s were then given two preliminary trials with a sequence of 64 figures in which difficulty of form discrimination was held constant at the least difficult level and in which no irrelevant dimensions were present. The *S*s' task was to identify each of the figures and to respond by moving the appropriate response switch. Upon completion of this training *S*s were read further instructions which informed them of changes in the shape of Forms A and B (discrimination difficulty) and the extra dimensions in the experimental sequences (irrelevant information). Following preliminary training *S* began Trial 1 with an experimental sequence appropriate to his group. Each trial consisted of 128 stimulus figures which were presented automatically for $\frac{1}{2}$ sec. followed immediately by a blank frame which served as the response period and lasted $5\frac{1}{2}$ sec. Each response and its latency were recorded by

E. After two trials *S* was dismissed until the next day. On each of the next two days *S* was given three trials (128 figures per trial) by means of an appropriate experimental film.

RESULTS

Since *S* performed under all four levels of discrimination and for eight practice trials, there were 32 error and 32 latency scores for each *S*.

Latency analysis.—In Fig. 2 the mean response latencies of the three groups with different amounts of irrelevant information (1, 2, and 3 pairs of irrelevant dimensions) are presented for Trials 1–8. It is seen that the mean latencies increase as the number of irrelevant dimensions are increased; at no time during the eight practice trials is there a reversal. An analysis of variance of the latency scores demonstrates that the effects of different amounts of irrelevant information (sometimes relevant) are significant beyond the .025 level (Table 1).

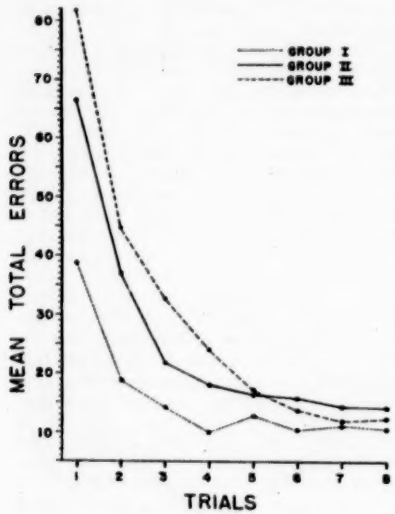

FIG. 3. Mean total errors as a function of number of irrelevant dimensions and stage of practice.

TABLE 1

ANALYSIS OF VARIANCE FOR THE LATENCIES OF RESPONSES

Source	*df*	*MS*	*F*
Between *Ss*	(29)		
Irrelevancy (I)	2	5,980.9	4.4*
Error (b)	27	1,374.7	
Within *Ss*	(930)		
Practice (P)	7	13,974.7	589.6***
Discrimination difficulty (D)	3	2,270.5	95.8***
P × D	21	17.7	
P × I	14	210.6	8.9***
Linear	(2)	1,050.8	5.9**
Quadratic	(2)	159.3	
D × I	6	24.8	1.0
P × D × I	42	3.5	
Error (w)	837	23.7	
Linear (P)	(27)	179.0	
Quadratic (P)	(27)	214.0	
Total	(959)		

* $P < .025$.
** $P < .01$.
*** $P < .001$.

With continued practice on the task *Ss* became more proficient in responding to the stimulus patterns ($P < .001$). The significant Practice × Irrelevancy interaction ($P < .001$) can be interpreted to mean that the different amounts of irrelevant information had different effects upon task performance at various stages of practice. Further analysis by means of the extended trend test (Grant, 1956) revealed a significant linear component in the interaction, i.e., the different irrelevancy curves actually have different slopes (Fig. 2). Comparison of the individual slopes with *t* tests showed that Groups I and II are significantly different from Group III, but that Groups I and II do not differ significantly. The significant irrelevancy factor indicates, then, that the over-all effect of sometimes-relevant irrelevant information is to reduce the efficiency of discrimination behavior, but the P × I interaction suggests that this between-groups effect tends to decrease with practice.

The results of the analysis also show that discrimination difficulty of the primary form dimension significantly affected the over-all task difficulty. The Discrimination-difficulty × Irrelevancy interaction is, however, insignificant.

Error analysis.—The mean total errors for the three irrelevancy groups are presented in Fig. 3. In general, the number of errors increased as the number of irrelevant dimensions is increased. Since the error scores were not normally distributed, separate nonparametric tests were employed to assess the effects of each major variable. When tested with the H statistic (Siegel, 1956), the irrelevancy variable was found to be significant at the .01 level. The results of the Friedman χ^2_r test (Siegel, 1956) showed that both practice and discrimination difficulty were also significant factors. The error data, thus, support the results of the latency analysis.

Discussion

The results indicate that increasing amounts of irrelevant information detrimentally influence complex discrimination performance. Since in the Archer (1954) study irrelevant information did not affect such performance, these results, together with the results of the preliminary study summarized in the introduction, strongly suggest that an important factor is that the irrelevant information be at other times relevant. Presumably, the decrement in performance results from an increased competition among responses. It seems, therefore, that both the amount and the nature of the irrelevant information is important.

The significant Practice × Irrelevancy interaction points up the fact that greater differences occurred among the irrelevancy conditions during the early practice trials. This raises the important question of the specific manner in which the influence of irrelevant information is reduced by practice. Do the differential effects produced by the irrelevant information decrease because the relevant discriminations are better learned, or must S learn to ignore the irrelevant information? If the first is the important consideration, practicing the relevant discriminations in the absence of irrelevant information should lead to as great a reduction in the detrimental effects of irrelevant information as practice with the irrelevant information present.

Summary

The study attempts to determine whether increasing amounts of irrelevant information that is relevant under other conditions detrimentally influences performance of a complex discrimination task. Two further questions asked were: (*a*) whether the effect of the irrelevant information is increased as the discrimination of the relevant information is made more difficult, and (*b*) whether the effect of the irrelevant information is reduced by practice. On a given trial *S*s were required to identify one of 16 complex geometrical patterns to which had been added either 1, 2, or 3 pairs of irrelevant dimensions. An analysis of variance of the response latencies showed that differences among the irrelevancy conditions were significant ($P = .025$). Both the discrimination and practice variables were highly significant, but only the Practice × Irrelevancy interaction was significant. The error data support the results of the latency analysis.

References

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Hodge, M. H. The influence of irrelevant stimuli upon complex visual discrimination. Unpublished doctoral dissertation, Univer. of Virginia, 1957.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received January 13, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

EFFECTS OF INSTRUCTIONS ON THE TRANSFER OF A CONDITIONED RESPONSE [1]

RICHARD H. LINDLEY

Carnegie Institute of Technology

Wickens (1938, 1939, 1943a, 1943b, 1948) conditioned *S*s to make extensor responses of their fingers and then reversed the position of their fingers in relation to the electrode. Most *S*s transferred with the antagonistic muscular response of flexion. Since antagonistic responses are reciprocally inhibited, the flexor responses could not have been part of the originally conditioned movement and thus could not have been conditioned during the original training.

The present study tested the hypothesis that *S*s instructed themselves to move away from the electrode during transfer. The hypothesis was tested by varying the type of initial instructions read to *S*s. It was predicted that one group would transfer toward the electrode and another group would transfer away from the electrode as a result of the instructions. The effects of instructions on the unconditioned response were also studied.

METHOD

The method for Groups 1, 2, and 3 will be described first. A fourth group, which was tested after the first three groups had been tested, will then be described.

Apparatus for Groups 1, 2, and 3.—The CS was a tone of 400 cps generated by a Jackson oscillator which *S* heard through earphones. The UCS was produced by an inductorium. The right hand rested palm down on a large brass electrode with the middle finger extended and resting between the rollers of two microswitches. The rollers touched the dorsal and volar surfaces of the finger respectively; shock could be delivered through either roller. A movement of the finger up or down 1/16 in. tripped the appropriate microswitch and flashed a light on a control panel in front of *E* showing the direction of finger movement.

The CS-UCS time interval was controlled by a cam and constant speed motor arrangement. Response latency was measured by an electric timer.

Procedure for Groups 1, 2, and 3.—The *S* was seated in a chair parallel to the conditioning board. The instructions were read and the arm was strapped in the conditioning board. The microswitches were adjusted about 2 in. proximal to the finger tip.

Discrete shocks of .2-sec. duration were delivered to the finger. Shock intensity was increased until *S* gave a fairly vigorous finger movement. Three tests with tone alone were given. If *S* responded, the tests were continued until no response occurred for three consecutive trials.

Avoidance conditioning was used. The CS was presented .5 sec. before the UCS, which was of .2 sec. duration. The CS overlapped the UCS completely. Any response to the CS with latency .5 sec. or less avoided the UCS on that trial.

The paired presentations of the CS and UCS were separated by 5- to 25-sec. intervals. A prearranged sequence of time intervals was used which was the same for all *S*s. After *S* reached criterion of four conditioned responses in five consecutive trials, the hand was turned palm up for the transfer test.

Three transfer tests with tone alone for .8 sec. were given. If *S* responded to the tone, the tone was terminated. No shock was delivered in the transfer test.

Six *S*s did not respond in the transfer test. Five reconditioning trials with palm down and three transfer tests with palm up were given to these *S*s. Five *S*s now transferred; these data are not reported separately.

Half the *S*s in each group were conditioned to extend and half were conditioned to flex. The *E* and *S* were in the same room with an opaque screen separating them.

[1] This paper is based on a dissertation submitted in 1957 in partial fulfillment of the requirements for the Ph.D. degree at the University of California. The author wishes to express his gratitude to Donald A. Riley for his encouragement and advice during the course of this research. Thanks are also due James L. McGaugh for his invaluable technical assistance.

Experimental design.—The instructions for Group 1 were comparable to those used by Wickens (1943b) and were designed to produce antagonistic responses in transfer. The *Ss* conditioned to extend were told to move their fingers up in the original conditioning. If the instruction to move up carried over to the transfer test, *Ss* would make the antagonistic muscular response. The instructions were suitably modified for *Ss* conditioned to flex.

The instructions for Group 2 were designed to produce a transfer response toward the electrode. The instructions emphasized that a specific muscular response (either extension or flexion) was to be conditioned. Furthermore, *E* had *S* demonstrate the muscular response both with the palm up and the palm down before conditioning started.

Hilgard and Marquis (1940) report that". . . some subjects have been found to misunderstand the demands of the experiment so completely as to leave their fingers on the electrodes throughout the shock" (p. 270). The hypothesis that the unconditioned response itself is determined by self-instructions was tested in Group 3. The instructions and transfer test for Group 3 were the same as for Group 1. In Group 3, *S* was deliberately misled as to the location of the electrode which delivered the shock so that, by following instructions, *S* would move toward the shock.

Apparatus and procedure for Group 4.—In Group 4, the middle finger of the right hand with palm down was flexed at a 90° angle over the edge of the conditioning board and rested between two microswitches. The electrodes were two wires taped to the volar surface of the finger, about 1½ in. from the finger tip. The initial instructions were comparable to those used for Groups 1 and 3. All *Ss* were conditioned to extend; in all other respects the conditioning procedure was the same as before. In the transfer test, the hand was not turned over but *Ss* were instructed to make the antagonistic response. The *S* was cautioned not to respond voluntarily.

Subjects.—The *Ss* were 78 students from the University of California. The data of 22 *Ss* who did not reach criterion in 35 trials and of 2 *Ss* who did not transfer are not reported here. The *Ss* were assigned randomly to Groups 1, 2, and 3, except that, after 12 *Ss* had been conditioned in Group 1, no more *Ss* were assigned to that group.

Results and Discussion

Original conditioning.—The overall median number of trials to reach criterion for Groups 1, 2, and 3 was 16.5 trials with a range of 5 to 40 trials. Table 1 presents the data of the number of *Ss* in each group who were above and below the overall median.

TABLE 1

Number of *Ss* in Each Group Who Were Below and Above the Over-All Median Number of Trials of Original Conditioning

Group	*Ss* Below Median	*Ss* Above Median	Total
1	4	8	12
2	5	11	16
3	13	3	16
Total	22	22	44

The *Ss* who moved toward the shock reached criterion in significantly fewer trials than those who moved away from it (χ^2 is 9.83, $P < .01$ with 2 *df*). In the shock-withdrawal groups, *S* lost contact with the electrode as soon as the finger movement started; in the shock-approach group, *S*'s finger remained in contact with the electrode until the movement was completed. Therefore, *Ss* in the approach group presumably received a shock of longer duration than *Ss* in the withdrawal groups. Assuming that a longer UCS is also a stronger UCS, the difference in speed of acquisition of the CR is explained since speed of acquisition is directly related to intensity of UCS (Spence, 1953).

The *Ss* responded to the shock on 95.10% of the trials on which it was delivered. The *Ss* instructed to withdraw from the electrode withdrew on every trial on which a movement was made, and *Ss* instructed to approach approached on every trial on which a movement was made. Thus in the present experiment the UCR to shock was controlled by instructions.

TABLE 2

DIRECTION OF TRANSFER AFTER THE POSITION OF THE HAND WAS REVERSED

Group	*Ss* Who Transferred With the Originally Conditioned Response	*Ss* Who Transferred With the Antagonistic Response	Total
1	4	8	12
2	11	5	16
3	3	13	16
Total	18	26	44

Transfer responses.—Table 2 presents the results for the direction of transfer responses. The direction of transfer was dependent on the initial instructions (χ^2 is 8.66, $P < .02$ with 2 *df*).

In Group 4, the 10 *Ss* who conditioned transferred with the antagonistic response of flexion (χ^2 corrected with Yates' correction is 8.10, $P < .01$ with 1 *df*). Instructions at the time of transfer determine the direction of transfer.

The transfer results clearly indicate that the direction of conditioned finger movement after hand reversal is a function of the instructions read to *Ss*. Since the UCR to shock was not a reflex, caution should be exercised in generalizing the results of this experiment. Conditioned reflex responses may not be controlled by instructions; further research is needed to answer this question.

These results do not contradict the response generalization explanation of antagonistic transfer proposed by Wickens (1943a). Rather, they may indicate the manner in which response generalization is accomplished, i.e., through self-instructions.

SUMMARY

The effects of instructions on the direction of transfer of a conditioned finger-withdrawal response and on the direction of movement to shock were investigated. The results showed that instructions were an important determiner of the responses. The implications of these results were discussed.

REFERENCES

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

WICKENS, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *J. exp. Psychol.*, 1938, **22**, 101–123.

WICKENS, D. D. The simultaneous transfer of conditioned excitation and conditioned inhibition. *J. exp. Psychol.*, 1939, **24**, 332–338.

WICKENS, D. D. Studies of response generalization in conditioning: I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, **33**, 221–227. (a)

WICKENS, D. D. Studies of response generalization: II. The comparative strength of the transferred and non-transferred responses. *J. exp. Psychol.*, 1943, **33**, 330–332. (b)

WICKENS, D. D. Stimulus identity as related to response specificity and response generalization. *J. exp. Psychol.*, 1948, **38**, 389–394.

(Received January 2, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

ADDITIVITY OF CUES AND TRANSFER IN DISCRIMINATION OF CONSONANT CLUSTERS[1]

FRANK RESTLE

Michigan State University

Examples of additivity of cues, found in the selective learning of rats (Blodgett, McCutchan, & Mathews, 1949; Eninger, 1952; Galanter & Shaw, 1954; Scharlock, 1955) and monkeys (Warren, 1953), have been analyzed by the writer in papers on a mathematical model of discrimination learning (Restle, 1955, 1957, 1958). The present study extends these findings and the analysis to human learning of simple two-choice discriminations.

In order to decide whether or not cues are additive, it is necessary to decide what behavioral manifestations will be taken to indicate additivity. Since learning data by their nature reflect changes in the organism, and the process of learning is complicated, there is some difficulty in selecting an index which can rationally be expected to reflect additivity of cues. The writer's approach to this problem has been to assume that in simple two-choice discrimination learning, the proportion of relevant cues determines the rate of learning. This implies that additivity of relevant cues will be manifested as additivity of learning rates, which can be estimated.

The mathematical model to be used states that discrimination learning involves conditioning relevant cues to the correct response and neutralizing,[2] or "adapting," the irrelevant cues. A neutralized cue has no direct effect on response, and other cues contribute toward the probability of a correct response or an error accordingly as they are conditioned to the correct or wrong response. The rates of conditioning and neutralization are assumed to be equal and to equal the proportion of relevant cues in the problem (Restle, 1955).

Such assumptions, suitably expressed in difference equations, imply a logistic-type learning function depending on a single parameter θ, the proportion of relevant cues. For a given class of *S*s, it is possible to estimate the θ characteristic of a problem from performance on that problem using Restle's Equation 8 (Restle, 1955).

To test additivity of cues, one constructs a problem with Cue *a* relevant, another problem with *b* relevant, and a third problem with both *a* and *b* relevant and redundant. The number of relevant cues in the third problem should be the same as the sum of the numbers relevant in the first two problems, if additivity is to hold. If the total number of cues is the same in all three problems, then the *proportions* of relevant cues will also add. The total number of cues may be held constant by ensuring that the same cues are present in all three problems; for example,

[1] This study was supported in part by the All University Research Fund, Michigan State University. Data were collected by James Richmond.

[2] The term "neutralize" is used in place of the word "adapt," which appears in theoretical papers on this subject (Restle, 1955, 1957, 1958). "Adapt" has been criticized as having other and better usage in experimental psychology. "Neutralize" has been used by D. LaBerge with a meaning very like that employed in the present connection.

in the first problem, Cue *a* is relevant, but Cue *b* is also present and irrelevant.

Additivity of Cues *a* and *b* means that they neither coalesce, nor compete for attention, nor interact in any way which would have an important effect on their attention values. It was judged that letters in a consonant cluster, especially if separated by an unchanging letter, might satisfy such conditions. The first purpose of the experiment was to test this hypothesis.

A second purpose of the experiment was to study transfer from the two-cue to a one-cue problem in this simple situation. The *S* learns a problem which can be solved using either of two cues, and then one of the cues becomes irrelevant, the other remaining relevant and with unchanged relationship to the correct answers. According to the theory, *S* must neutralize the newly irrelevant cues to master this new problem. From the first part of the experiment we have estimates of the proportions of cues still relevant and of those newly scrambled, and the theory asserts that the rate of neutralizing of scrambled cues will equal the proportion of cues still relevant.

During the initial learning, there were some cues present other than Cues *a* and *b*. These "background" cues, *c*, which are irrelevant throughout the experiment, were neutralized during original learning. It appears that neutralization cannot be adequately described as "learning to ignore" cues, for a neutralized cue remains ineffective only so long as the relevant cues remain present and relevant—it seems that the process may be described as "discriminatory," and that an irrelevant cue is neutralized *with respect to* some relevant cue (Hammer, 1955; Restle, 1958). It this is so, scrambling part of the relevant cues may release some of the background (*c*) cues from their neutral state, and they may regain their disruptive influence. Using the mathematical model, and estimates of the relative weight of *a*, *b*, and *c* cues obtained from original learning, it may be possible to estimate the proportion of *c* cues which become deneutralized.

Method

Subjects.—The *Ss* were 80 students in elementary psychology at Michigan State University who receive course credit for taking part in experiments. They were tested individually, being assigned to groups as they appeared, in an irregular, prearranged order.

Apparatus.—A standard memory drum was modified to take a continuous tape with 48 stimuli and 48 correct answers. On each trial the stimulus was presented for 4 sec., the correct answer followed for 1 sec., and an intertrial blank lasted 1 sec.

Procedure and stimuli.—The *Ss* were told that they would see two types of consonant clusters, Type 1 and Type 2, and that they were to learn to classify them correctly. They were told that correct classification depended on the letters present and nothing else. The *Ss'* answers were given orally during the stimulus presentation, knowledge of results being automatic since the correct answer replaced the stimulus in the aperture. Late or omitted responses were recorded separately and considered as one-half error each. After each block of five trials, the drum was stopped manually in the blank phase and *S* was asked, "What hypotheses are you working on now?" and "Can you give further details?" Answers to these questions were recorded by *E*.

Three problems were prepared, all involving the same letters. Every cluster in every problem had the letters F, Q, and J in the first, third, and fifth positions, respectively. The second letter was either *X* or *Y*, and cues from this position are designated as Cue *a*. The fourth letter was either W or V, and cues from this position are designated as Cue *b*. In Problem A, the *a* cues are relevant and the *b* cues are varied randomly. In Problem B, the *b* cues are relevant and the *a* cues are varied randomly. In Problem A + B, both *a* and *b* cues are relevant. Type 1 and Type 2 stimuli appearing in each problem are shown in Table 1.

There were four experimental groups of 20 *Ss* each. Group I had 48 trials on Problem A,

TABLE 1

TYPE 1 AND TYPE 2 STIMULI APPEARING IN EACH PROBLEM

Problem	Type 1 Stimuli	Type 2 Stimuli
A	FXQWJ FXQVJ	FYQWJ FYQVJ
B	FXQWJ FYQWJ	FXQVJ FYQVJ
A + B	FXQWJ	FYQVJ

and Group II had 48 trials on Problem B. These groups gave data from which estimates of the proportional weights of *a* and *b* cues could be estimated. Groups III and IV started with 24 trials on Problem A + B, and then Group III had 24 trials on Problem A and Group IV had 24 trials on Problem B. The initial learning of these groups yielded data to test additivity of cues, and their transfer performance was analyzed with respect to the nature of neutralization of *c* cues, as discussed above.

RESULTS

Problem A + B was learned faster than either Problem A or Problem B, and Problems A and B were about equally difficult. Scrambling a cue caused a sharp drop in performance, from which recovery was gradual but steady. Figure 1 shows the group average learning curves.

For statistical analysis, the performance of each *S* was summarized by the percentage of correct responses on Trials 2–24, and Trials 26–48. Trial 1 was discarded because it could not reflect learning, and Trial 25 was discarded to make the various groups comparable, and because *S*s in Groups III and IV could not know the correct answer on that trial. Means and *SD*'s of these scores are shown in Table 2.

The data were nearly enough normal to permit use of the *t* test. Comparison of performance on Trials 2–24 between Group I and combined Groups III and IV indicated that Problem A + B is easier than Problem A. A comparison of Group II with combined Groups III and IV indicated that Problem A + B is also easier than Problem B. Both differences were at the .001 level of significance. Using all the data on Groups I and II, it was found that Problems A and B did not differ significantly and that performance differences were small. Raw transfer effect, measured by comparing the transfer performance (Trials 26–48) of Group III with Trials 2–24 of Group I, and transfer of Group IV with Trials 2–24 of Group II, was very small and not significant. Lawrence (1952) and Baker and Osgood (1954) have found cases in which transfer from an easy to a hard dis-

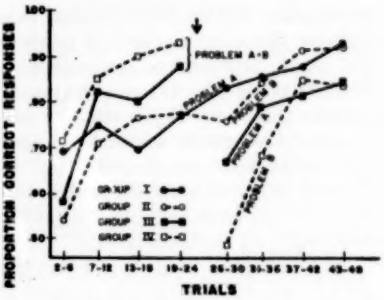

FIG. 1. Mean proportion of correct responses as a function of training trials, for Groups I, II, III, and IV. Arrow indicates beginning of transfer of training for Groups III and IV. Trials 1 and 25 are deleted.

TABLE 2

MEANS AND *SD*'S OF PROPORTIONS OF CORRECT RESPONSES ON TRIALS 2–24 AND 26–48

Group	Problems	Trials 2–24		Trials 26–48	
		Mean	*SD*	Mean	*SD*
I	A → A	.719	.170	.870	.159
II	B → B	.676	.176	.859	.142
III	A + B → A	.783	.145	.769	.191
IV	A + B → B	.853	.121	.686	.166

crimination is even more efficient than training the same number of trials on the hard problem. Comparison of the late trials (26–48) of the transfer groups (III and IV) with the late trials of the groups trained throughout on the hard problems (I and II) showed that training throughout on the hard problem is more efficient in this study, at the .001 level.

An anomalous result is that Groups III and IV, both of which worked on Problem A + B on Trials 2–24, were not equally successful. Group III made more errors than Group IV, and the difference is significant at about the .01 level. A recheck of subject sampling procedures, hours of day or days of the week on which *S*s of the two groups were run, and experimental procedures, failed to find an explanation for this difference. Group IV did not maintain its superiority during transfer despite the nearly equal difficulty of the two test problems.

Since the various comparisons reported above are not all mutually independent, they should be taken as descriptive. The hypothesis of additivity of cues, and the analysis of the nature of neutralization of cues, in no case depend on the statistical tests.

Additivity of cues.—If Cues *a* and *b* are additive, then θ_A (estimated from the performance of Group I) plus θ_B (estimated from the performance of Group II) should equal θ_{A+B} (estimated from the original learning of Groups III and IV combined). The prediction is exact for a single *S*, but since the three estimates cannot be obtained on a single *S*, we must settle for the central tendency of groups. The median interval, used as the index of central tendency in this analysis, is invariant over monotone transformations; or, more exactly, the transformation of the median interval is the median interval of the transformed scores. The *mean* of transformed values is not the transformation of the mean in the present case, and the discrepancy is large. Two estimation methods were employed on the data of each *S*, one based on total errors starting with Trial 2, the other based on weighted errors where each error is weighted by the number of trials of training which precede it. The second method gives emphasis to late errors, and has some of the advantages of a trials-to-criterion score, which could not be used in this study. The two methods gave comparable estimates, and the average of the two was used in further computations.

The median interval used here is the interval between the estimated values of θ of *S*s who ranked tenth and eleventh in the group of 20. The median interval for Group I on Problem A was (.099, .115). The median interval for Group II on Problem B was (.109, .122). Since the most we know about the "true" median of a group is that it is somewhere in the median interval, the closest prediction possible for Problem A + B is that the median for Groups III and IV should fall in the interval (.099 + .109, .115 + .122) = (.218, .237). Because of ties, the combined Groups III and IV had a point median value of .230, which is within the predicted interval. Of the 40 *S*s, 18 fell below the predicted median interval, 17 were above it, and five scores fell within the interval. Thus, the hypothesis of additivity fits the data exactly.

Transfer of training.—The performance on Problems A or B following learning of A + B was analyzed to find out whether the *c* cues, those irrelevant throughout the experiment, become deneutralized when transfer began. For the sake of simplicity it is assumed that $\theta_A = \theta_B = .115$, and θ_C, the proportion of *c* cues, equals .770. These assumptions are quite close to the estimates obtained. By setting $\theta_A = \theta_B$ one assumes that transfer from A + B to A will be the same as transfer from A + B to B. We recall that Groups III and IV did not differ significantly in proportion of correct responses in transfer.

Consider the case of transfer from

Problem A + B to Problem A (with the above assumptions, exactly the same result will obtain for transfer from A + B to B): the error-producing cues in transfer will be the b cues scrambled plus any deneutralized c cues. The proportional weight of these cues is $\theta_B + d\theta_C$, where d is the proportion of c cues which become deneutralized. These error-producing cues become adapted, according to the theory, with a probability θ_A on each trial of transfer. The b and c cues, being irrelevant in a two-choice problem, have a probability of $\frac{1}{2}$ of producing an error, on the average.

From the argument advanced earlier (Restle, 1955), the probability of an error on Trial j of transfer is

$$q_j = \frac{.5(\theta_B + d\theta_C)(1 - \theta_A)^{j-1}}{\theta_A + (\theta_B + d\theta_C)(1 - \theta_A)^{j-1}}$$

$$= \frac{.5(.115 + .770d)(.885)^{j-1}}{.115 + (.115 + .770d)(.885)^{j-1}}$$

An approximate sum of this equation for $j = 2$ to 24 can be obtained. This corresponds to the expected errors on the transfer Trials 26–48, as a function of d. Mean and median errors on Trials 26–48 for Groups III and IV were about 6.0. Setting the expected errors equal to 6.0 and solving for d, we obtain $d = .52$, indicating that about half the background c cues became deneutralized when one of the two relevant cues was made irrelevant.

The above computations are hardly precise. It is assumed that the learning of Problem A + B was complete, and though more than half the *S*s in Groups III and IV had reached a criterion of 10 consecutive correct answers by Trial 24, a number had not solved the problem. In addition, exact values of the parameters θ_A and θ_B were assumed, and they were assumed equal. If, despite these possible sources of error and the great variability of the transfer data, we take the result seriously, it suggests that scrambling half the relevant cues in a problem results in deneutralization of half the previously neutral irrelevant cues.

Discussion

The present experiment yielded a good quantitative check of additivity of cues arising from separate letters in a consonant cluster. The case for expecting additivity of cues was very strong, so that a failure of the prediction would have been serious. The use of discrete, separated cues, both visible during the usual inspection time, was intended to avoid coalescing or competition; and identically the same cues were available in Problems A, B, and A + B, so that excuses having to do with configural changes due to removing stimuli could not be supported. However, since the two cues turned out to be about equally strong, the result throws no light on Babb's (1957) failure to find additivity of unequal cues.

Another way of looking at the additivity result, suggested by the work of Gulliksen and Wolfle (1938), Bitterman and his associates (Turbeville, Calvin, & Bitterman, 1952), and Estes,[3] is to notice that in Problem A + B only two configurations or stimulus patterns are used, whereas in Problems A and B, there are four configurations. If the separate configurations are learned separately, at a fixed rate, learning should go twice as fast in Problem A + B as in Problem A or Problem B. Inspection of Fig. 1 shows that Groups III and IV reached about 92% correct in 24 trials, and Groups I and II reached about the same level of performance in 48 trials. This is in accord with the configural hypothesis. Furthermore, Table 2 indicates that Groups III and IV had about 81.3% correct responses during Trials 2–24, while Groups I and II had about 78.1% correct on Trials 2–48. This indicates that the level of the two learning curves was about comparable, again in accord with the configural hypothesis. Analysis of the transfer of training data according to this hypothesis is beyond the scope of the present report.

[3] W. K. Estes. Personal communication, 1957.

The transfer of training performance is consistent with the idea that irrelevant cues are "neutralized" relative to the relevant cues present, and recover if the relevant cues are removed or scrambled. This would explain Hammer's (1955) failure to find a transfer of neutralization, since that experiment did not carry any relevant cues from one problem to the next. The idea that cues are neutralized by means of an observing response (Wyckoff, 1952) would also lead to the conclusion that a cue remains neutral only so long as the relevant cue remains relevant.

Summary

Four groups of 20 human *S*s each were trained on problems with one relevant and one irrelevant cue, or with both cues relevant. The stimuli were consonant clusters, and the responses were verbal. The problem with two relevant cues was easier than the others, and theoretical analysis indicated that the cues were additive. When *S*s were shifted from the two-cue to the more difficult one-cue problem, only slight positive transfer was found. The quantitative results suggested that when a hitherto relevant cue became irrelevant, some previously neutralized irrelevant cues regained their function. A second quantitative analysis, based on the number of configurations in the problems, was discussed briefly and found consistent with the additivity-of-cues data.

References

Babb, H. Transfer from a stimulus complex to differentially discriminable components. *J. comp. physiol. Psychol.*, 1957, **50**, 288–291.

Baker, R. A., & Osgood, S. W. Discrimination transfer along a pitch continuum. *J. exp. Psychol.*, 1954, **48**, 241–246.

Blodgett, H. C., McCutchan, K., & Mathews, R. Spatial learning in the T-maze: The influence of direction, turn, and food location. *J. exp. Psychol.*, 1949, **39**, 800–809.

Eninger, M. U. Habit summation in a selective learning problem. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

Galanter, E. H., & Shaw, W. A. "Cue" vs. "reactive inhibition" in place and response learning. *J. comp. physiol. Psychol.*, 1954, **47**, 395–398.

Gulliksen, H., & Wolfle, D. L. A theory of learning and transfer: I. *Psychometrika*, 1938, **3**, 127–149.

Hammer, M. The role of irrelevant stimuli in human discrimination learning. *J. exp. Psychol.*, 1955, **50**, 47–50.

Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Restle, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64**, 217–228.

Restle, F. Toward a quantitative description of learning-set data. *Psychol. Rev.*, 1958, **65**, 77–91.

Scharlock, D. P. The role of extramaze cues in place and response learning. *J. exp. Psychol.*, 1955, **50**, 249–254.

Turbeville, J. R., Calvin, A. D., & Bitterman, M. E. Relational and configurational learning in the rat. *Amer. J. Psychol.*, 1952, **65**, 424–433.

Warren, J. M. Additivity of cues in visual pattern discriminations by monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 484–486.

Wyckoff, L. B. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received January 21, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

DISCRIMINATION IN THE SQUIRREL MONKEY AS A FUNCTION OF DEPRIVATION AND PROBLEM DIFFICULTY [1]

RAYMOND C. MILES

Montana State College

Most investigations demonstrating that performance improves with increasing deprivation have used learning criteria related to the general activity level of *S*s. These measurements include time scores for traversing a maze or runway (e.g., Barry, 1958; Lewis & Cotton, 1957), subsequent trials to extinction (e.g., Cautela, 1956), rate of responding (e.g., Dinsmoor, 1952), and responses emitted during extinction of lever pressing and panel pushing (e.g., Miles, 1956). Because increasing deprivation is correlated with an increase in activity level, the precise relationship of deprivation to performance of a learned repertoire is difficult to evaluate when a rate measure is involved (Miles, 1958; Siegel & Steinberg, 1949).

There are, however, several experiments in which the effect of drive upon performance was assessed with a measure—number of correct responses in a two-choice discrimination—relatively free from the direct effects of bodily activity. The results of the latter investigations do not lead toward any unified conclusion, although a perceptible majority report no influence on performance due to increased drive.

Meyer (1951) found that performance of rhesus monkeys on object discriminations was not influenced by degree of deprivation. Teel (1952) trained albino rats on a position habit in a T maze and found no effect of drive on performance. Comparable results were obtained by Hillman, Hunter, and Kimble (1953), who trained rats on a 10-unit elevated T maze; although time scores revealed reliable differences in performance, deprivation had no effect on error scores. Because of this independence between time and error scores, the authors advanced a thesis analogous to the view taken in the present investigation: "Measures involving a time dimension are more responsive to motivational changes than measures, like number of errors, which are independent of time" (Hillman, et al., 1953, p. 89). Experiments by Anderson (1941) and by Bruce (1938), using different measures of maze-learning proficiency, also support this thesis.

Data from other two-choice discrimination investigations do indicate that increased drive facilitates performance. Buchwald and Yamaguchi (1955) tested rats on a simple T maze and found better performance under the condition of high water deprivation. Since numbers of reinforcements during original learning were not entirely equalized, they believed that the most clear-cut results involved a subsequent reversal learning situation. This added complication, involving unlearning a previous position habit, makes a direct comparison with other experiments rather difficult. An experiment in which rats were trained on a black-white discrimination by Eisman, Asimow, and Maltzman (1956) also found evidence that drive improves discrimination performance. It is of significance to note that performances after deprivations of 4 and 22 hr. did not differ from each other, but that performance after a relatively severe deprivation of 46 hr. was reliably superior to performance under both lower drive conditions.

The inconclusive status of the problem is not surprising when one considers that species, apparatus, designs, types and lengths of deprivations, learning criteria, and other

[1] This investigation was supported in part by the Graduate School of the University of Wisconsin, and in part by the National Institute of Neurological Diseases and Blindness, National Institutes of Health, through Grant B-732 to Clinton N. Woolsey.

conditions differed among the various experiments. It does appear warranted, however, to generalize that the facilitating effect of deprivation on performance is relatively easy to demonstrate with rate measures; and, in contrast, is much more difficult to demonstrate in two-choice discrimination situations where number of correct responses is the measure of learning.

The plan of the present investigation is essentially to duplicate and to extend the Meyer (1951) experiment in which rhesus monkeys were tested on two-choice object discriminations. In Meyer's carefully designed experiment, the proficiency of the rhesus monkey in solving discrimination reversals was found to be independent of degree of food deprivation when tests were given after 1, 23, and 47 hr. The generality of these findings may be somewhat restricted by a few minor conditions which may have decreased the importance of deprivation. A preferred food (one not included in the usual diet) served as the reinforcement for testing, and, in keeping with established testing procedure, *Ss* were prefed a small quantity of the test incentive before each test session. The *Ss* had had some experience with discrimination learning and were able to solve the discrimination reversal problems easily. Perhaps a more difficult task is required in order to disclose any performance difference due to deprivation.

An attempt was made in the present experiment to control all conditions which might diminish the contribution of the deprivation parameter. The squirrel monkey *Ss* were naive; the test incentive, which was included in the daily diet, was not prefed to *Ss* prior to test sessions; and two levels of discrimination difficulty were included.

METHOD

Subjects.—The *Ss* were four naive, adult, male squirrel monkeys, *Saimiri sciureus;* they were housed and tested at the University of Wisconsin Primate Laboratory. Their daily diet consisted of perishable fruits and vegetables, cooked eggs, bread, and a vitamin supplement. A small handful of grapes, the food used as the incentive during testing, was also included in the daily diet. All food was removed at the termination of a 1-hr. eating period, with the end of this period marking the beginning of deprivation.

Apparatus.—Testing was conducted in a modified Wisconsin General Test Apparatus, which was constructed on a smaller scale and adapted for visual discrimination. The equipment consisted of a plywood box, $40 \times 22 \times 15$ in., divided into two compartments by a series of vertical bars. An opaque screen in front of the bars could be raised or lowered by *E*. One compartment housed *S* during the test period; the other section, the test compartment, was equipped with two overhead light sources (150-w. projectors), a movable test tray, and two test objects. The test tray and the floor of the test compartment were painted flat black, and a 1½-in. glossy white square was centered around each of two small food wells. These food wells were spaced 12 in. apart and were located 1 in. from the front edge of the test tray. Circular test objects, 2¼ in. in diameter and ½ in. high, covered the food wells.

A fine-grained white paper, replaceable when soiled, was glued to the top surface of the test objects. Two similar reflecting surfaces were on the floor of the test compartment, so placed that the two overhead light sources projected a circular spot of light upon them. When the test tray was moved to the forward position, it passed over these diffusing surfaces and the light then shone directly on the test objects. Four neutral density filters were used to vary luminance. These rested in filter holders at right angles to the light sources and intercepted the light rays intended for the reflecting surfaces.

Procedure.—Since it was essential that *Ss* adjust to the once-daily eating schedule, this routine began two weeks before the beginning of the experiment. As an introduction to testing, *Ss* were first trained to push a single block from a food well in order

to obtain a food reward. Pretraining was continued for 25 trials.

A test trial consisted of the following sequence: The test tray was withdrawn while the opaque screen was being lowered in front of *S*, and a small quarter of grape was placed in the designated food well. Preparation was completed by placing the test objects on the food wells and inserting into the filter holders the appropriate neutral density filter combinations. While the test tray remained in the withdrawn position, the opaque screen was raised. The light sources thus illuminated the reflecting areas on the test compartment floor until, after a 1-sec. pause, the tray was advanced and the circular test objects received the illumination. The *S* could then displace either test object and uncover a food well. A correct effort was immediately reinforced with the food reward, whereas an incorrect response was followed by quick retraction of the tray.

The purpose of the 1-sec. pause between raising of screen and advancing of tray was to give *S* a controlled viewing period. In order to preserve objectivity, *E* did not observe *S* during the pause and timed the interval by slowly saying "One." The right-left position of the correct luminance changed from trial to trial according to a Gellermann series. Other controls included inserting filters into both holders on every trial to prevent differential noise cues, smearing both food wells with the reinforcement as a control for odor cues, and observing through one-way glass. The *S*'s response was recorded after each trial.

The discrimination problems included two levels of difficulty. The easier discrimination required *S* to differentiate between densities of 1.50 and .05. For the difficult discrimination, the difference to be discriminated was .70 vs. .40. These filters intercepted projected light rays from the 150-w., 2 × 2-in. slide projector focused on infinity; the lenses were 16 in. from the reflecting surfaces of the test objects. To maximize positive transfer effects, the brighter luminance was rewarded under both conditions.

Each experimental session consisted of 50 trials, with blocks of 5 easy discrimination trials alternating with 5 difficult discriminations until *S* had received 25 trials on each level. Testing was continued until each *S* had received a total of 1950 trials.

There were three lengths of deprivation period, 1, 5, and 20 hr., and each *S* was tested under all three deprivation conditions before any one again occurred. The order of presentation of the conditions was randomized by permuting every sequential combination of the three deprivation intervals and then using a random selection of these combinations.

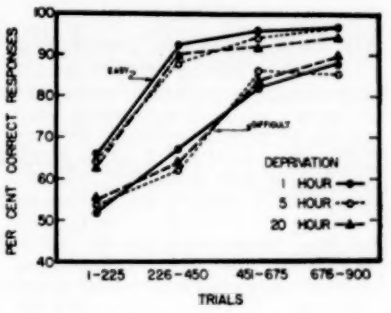

FIG. 1. Mean percentage correct responses as a function of trials and deprivation conditions. Plotted against successive 225-trial blocks on each difficulty level are individual points representing the four *S*s' mean percentage of correct responses for the 75 trials under each one of the three deprivation conditions.

After completion of approximately one-third of the investigation, *E* noted that the grape skins frequently were not eaten by *S* and were left on the removable floor of the restraining compartment. Since number of grape skins could serve as an independent measure of drive during actual testing, all skins remaining after a test session were counted and the number tabulated.

RESULTS

The major findings of the experiment are summarized in Fig. 1. This illustration shows percentage of correct responses as a function of deprivation for both the easy and the difficult discriminations throughout the course of the experiment. The lack of divergence in the learning curves for the three deprivation conditions indicates that degree of deprivation had no appreciable influence on performance at either level of discrimination difficulty.

Data based on total number of correct responses satisfied the condition of homogeneity of variance (Bartlett's test), and the results of

TABLE 1

ANALYSIS OF VARIANCE OF TOTAL NUMBER OF CORRECT RESPONSES

Source	*df*	*MS*	*F*
Subjects (*S*)	3	514	11.44*
Difficulty (Di)	1	18,370	408.22**
Deprivation (De)	2	140	
S × Di	3	40	
S × De	6	36	
Di × De	2	25	
S × Di × De	6	45	

* $P < .01$.
** $P < .001$.

an analysis of variance are presented in Table 1. In accord with expectation, the effects due to individual monkeys and to problem difficulty were highly reliable. Of more importance to the major experimental question was the failure to find any significant differences due to degree of deprivation.

Figure 2 shows the average number of grape skins left on the restraining compartment floor as a function of deprivation interval. This proved to be an excellent independent measure of deprivation during test sessions; as would be expected, the average number of grape skins decreased with increasing length of deprivation. Because Bartlett's test indicated nonhomogeneity of variance, these data were analyzed with the Friedman (1937) two-way analysis of variance. With the exception of the difference between the 5-hr. and the 20-hr. conditions ($P < .05$), all differences were significant beyond the .01 level.

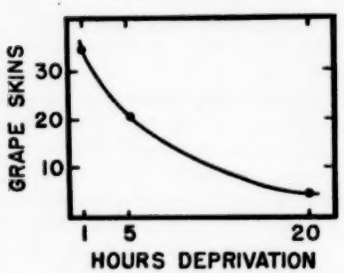

FIG. 2. Unconsumed grape skins as a function of deprivation. Each point represents the median number of grape skins remaining on the restraining compartment floor after the given length of deprivation.

DISCUSSION

The results of this investigation are consistent with those of Meyer's (1951) discrimination reversal experiment, indicating that deprivation within the range tested is a variable of negligible importance in two-choice discrimination problems. Throughout this entire investigation, regardless of problem difficulty or stage of learning, the squirrel monkeys showed no discernible performance difference as a function of deprivation. The fact that the number of grape skins remaining on the restraining compartment floor was inversely related to length of deprivation indicated an actual difference in drive level during the test sessions.

Predictions from Hull's (1952) theory and from Spence's (1956) modification of Hull's theory were not substantiated. Since all other determiners of reaction potential were constant, and since the test situation was a free-responding one with a decided majority of responses occurring to the reinforced stimulus, performance should have been primarily influenced by the deprivation variable. In addition, the fact that more food was consumed by *S*s under higher drive conditions indicates that there were differences in incentive motivation, *K* (amount of reward), which also should have had a facilitating effect upon performance.

Further research is required in order to determine how dependent the data of the present experiment are upon the type of species, the general testing procedure, and the nature of the discrimination problem. It appears that, as a drive becomes stronger, there is undoubtedly an increase in strength of the

consummatory response related to that particular drive state, as well as an increase in general bodily activity. Although these reactions have frequently been reflected by rate and magnitude measures of learning, it is much more difficult to demonstrate the influence of deprivation on number of correct responses in a choice situation.

The consistency between results of the two primate investigations should merit a general conclusion. Meyer's (1951) investigation used relatively advanced primates with some test experience and tested them on object quality discriminations immediately followed by reversal trials. The *Ss* for the present experiment were more primitive primates with no discrimination learning experience; they were tested on both easy and difficult brightness discriminations. The *Ss* in both experiments received hundreds of trials and showed no improvement in performance as a function of the deprivation conditions investigated. Food deprivation within the range tested is apparently an insignificant variable in discrimination learning by monkeys.

Summary

The concern of the present experiment was with the influence of deprivation on discrimination performance. Four naive squirrel monkeys, *Saimiri sciureus*, were tested on an easy and a difficult brightness discrimination; *Ss* were rewarded with a quarter of grape for selecting the brighter of two luminances. Length of food deprivation prior to the test sessions was either 1, 5, or 20 hr. Grape skins not consumed by *S* were counted after each test session, providing an independent measure of hunger during testing, and the number proved to be inversely related to deprivation. Measures of proficiency in discrimination learning were not influenced by degree of hunger. It was concluded that performance on a two-choice discrimination is relatively independent of drive magnitude.

References

Anderson, E. The externalization of drive: III. Maze learning by non-rewarded and satiated rats. *J. genet. Psychol.*, 1941, **59**, 397–426.

Barry, H., III. Effects of strength of drive on learning and on extinction. *J. exp. Psychol.*, 1958, **55**, 473–481.

Bruce, R. H. The effect of lessening drive upon performance by white rats in the maze. *J. comp. Psychol.*, 1938, **25**, 225–248.

Buchwald, A. M., & Yamaguchi, H. G. The effect of change in drive level on habit reversal. *J. exp. Psychol.*, 1955, **50**, 265–268.

Cautela, J. R. Experimental extinction and drive during extinction in a discrimination habit. *J. exp. Psychol.*, 1956, **51**, 299–302.

Dinsmoor, J. A. The effect of hunger on discrimination responding. *J. abnorm. soc. Psychol.*, 1952, **47**, 67–72.

Eisman, E., Asimow, A., & Maltzman, I. Habit strength as a function of drive in a brightness discrimination problem. *J. exp. Psychol.*, 1956, **52**, 58–64.

Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statis. Ass.*, 1937, **32**, 675–701.

Hillman, B., Hunter, W. S., & Kimble, G. A. The effect of drive level on the maze performance of the white rat. *J. comp. physiol. Psychol.*, 1953, **46**, 87–89.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Lewis, D. J., & Cotton, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, **50**, 189–194.

Meyer, D. R. Food deprivation and discrimination reversal learning of monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.

Miles, R. C. The relative effectiveness of secondary reinforcers throughout deprivation and habit-strength parameters. *J. comp. physiol. Psychol.*, 1956, **49**, 126–130.

Miles, R. C. The effect of an irrelevant motive on learning. *J. comp. physiol. Psychol.*, 1958, **51**, 258–261.

Siegel, P. S., & Steinberg, M. Activity level as a function of hunger. *J. comp. physiol. Psychol.*, 1949, **42**, 413–416.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Teel, K. S. Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, 1952, **45**, 188–191.

(Received February 3, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

EFFECT OF PRELIMINARY TRIALS ON RATE OF CONDITIONING IN A SIMPLE PREDICTION SITUATION [1]

DAVID LaBERGE [2]

Indiana University

The purpose of this study is to carry out an experimental test of a new assumption in statistical learning theory (LaBerge, in press). According to current stimulus sampling models (Burke & Estes, 1957; Bush & Mosteller, 1955; Restle, 1957), each stimulus element in an experimental learning situation is assumed to be connected to one of the response alternatives, A_i, $i = 1, 2, \ldots r$, which are observed or recorded by *E*. For example, in two-choice situations A_1 and A_2 may designate left and right key presses or left and right turns in a T maze. The present theoretical modification, on the other hand, assumes that, at least initially, some of the elements are connected to none of the observed or recorded responses, but rather are connected to unobserved or unrecorded responses such as postural adjustments, orienting responses, covert verbalizations, etc. In addition, it is assumed that, on a given trial, the organism continues sampling and responding until one of the observed responses occurs. Since the termination of a trial is contingent upon an occurrence of an observed response, the probability of a particular observed response on a trial depends only upon the distribution within the set of elements in the sample connected to observed responses. Elements connected to responses other than the observed alternatives are ineffective, or *neutral*, in determining which observed response shall occur. It is assumed that after the reinforcing event occurs, all neutral elements in the terminal sample become connected to one of the observed responses in the same manner as do non-neutral elements. Over a series of trials in a constant stimulus situation, then, the set of neutral elements should progressively become connected to the observed responses until the neutral elements are depleted and unobserved responses cease occurring. With a sufficient number of trials, therefore, the present formulation becomes identical with that of other models (Burke & Estes, 1957; Bush & Mosteller, 1955; Restle, 1957). Differences in theoretical predictions are expected to occur in connection with early trials when the proportion of neutral elements is large, and in connection with trials when new stimulus elements are presented, as in transfer studies.

Although the formal model is presented elsewhere (LaBerge, in press), its salient features may be summarized here as follows. The signal presented at the beginning of a trial is represented schematically in Fig. 1. For the two-choice situation the population of stimulus elements is partitioned into three sets:

C_0 = the set of neutral elements
C_1 = the set of elements connected to response A_1
C_2 = the set of elements connected to response A_2.

[1] The author is deeply indebted to C. J. Burke and W. K. Estes for their generous advice and criticism. Lee Roy Beach assisted with the preparation of the experiment and the analysis of data. This research was aided by a grant from the Graduate School of Indiana University.

[2] Now at the University of Minnesota.

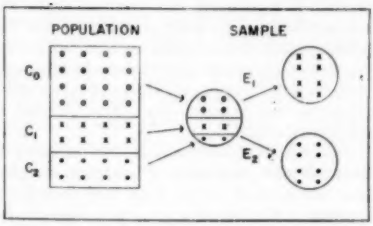

FIG. 1. Representation of random sampling of stimulus elements from a signal population, with the change in status of sampled elements due to an E_1 or E_2 reinforcing event.

Letting c refer to the number of elements in the respective sets, we may write the probability of an A_1 response on trial n as

$$p_{1,n} = \frac{c_{1,n}}{c_{1,n} + c_{2,n}} \qquad [1]$$

and the proportion of neutral elements on trial n as

$$u_n = \frac{c_{0,n}}{c_{0,n} + c_{1,n} + c_{2,n}} \qquad [2]$$

It is assumed here that the sampling probability, θ, is the same over all three sets of elements; it is further assumed that when an E_1 event occurs, the sampled elements become connected to A_1, and when an E_2 event occurs, the sampled elements become connected to A_2. Solution of the resulting difference equation[3] yields the mean learning function, which is expressed as the probability of an A_1 response in terms of ordinal trial number and parameters:

$$p_{1,n} = \pi - (1 - u_1)(\pi - p_{1,1}) \frac{(1 - \theta)^{n-1}}{1 - u_1(1 - \theta)^{n-1}} \qquad [3]$$

where $p_{1,1}$ is the initial probability and π is the probability of an E_1 event. When u_1 is zero, that is, when every element is initially connected to either A_1 or A_2, Equation 3 reduces to the Burke-Estes (1957) equation

$$p_{1,n} = \pi - (\pi - p_{1,1})(1 - \theta)^{n-1} \qquad [4]$$

as do Restle's (1957) function and Bush and Mosteller's (1955) linear model in comparable special cases. The present model agrees with the other three models in asymptotic prediction, for as n increases, Equations 3 and 4 approach the same value, π.

[3] The derivation requires the assumption that the population and sample sets be large in the same sense as do Burke and Estes (1957).

The denominator in Equation 3 contains the expression for the proportion of neutral elements as a function of trials:

$$u_n = u_1(1 - \theta)^{n-1} \qquad [5]$$

As n increases, the proportion of neutral elements decreases in a negatively accelerated manner, approaching zero asymptotically.

One prediction which may be derived from the present model is that the rate of learning will vary with the proportion of neutral elements in the population. This follows from an examination of Equation 3. It is clear that when the proportion of neutral elements, u_1, is near unity, the A_1 probability is near the asymptotic value, π, even after only one trial of training. When u_1 is zero, the A_1 probability after the first trial is nearer the initial value, $p_{1,1}$; and it can be shown rigorously that as u_1 increases, then for any given $n > 1$, $p_{1,n}$ will increase monotonically. Thus, as the proportion of neutral elements is increased, the rate of approach to asymptote is expected to increase.

A test of this prediction is carried out in the present study using a simple two-choice prediction situation in which the probability, π, of an E_1 event is .90, and the probability, $1 - \pi$, of an E_2 event is .10. In order to vary the proportion of neutral elements present at the beginning of this condition, four groups of *Ss* are given 0, 20, 60, or 200 preliminary trials under a 50:50 reinforcement condition. Those *Ss* having a greater number of preliminary trials are expected to have fewer neutral elements present when they are shifted to the 90:10 condition, and therefore they should show a slower approach to asymptote. Comparisons of the slopes of the learning curves after the shift to the 90:10 condition should provide a qualitative test of the prediction.

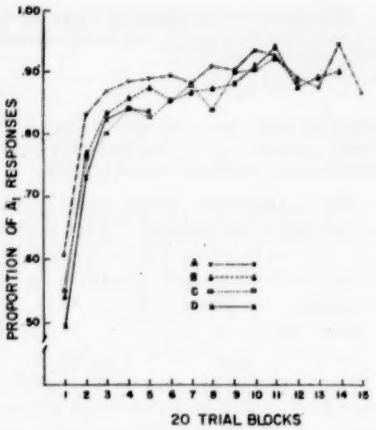

FIG. 2. Obtained conditioning curves under 90:10 reinforcement, after varying amounts of preliminary trials under 50:50 reinforcement. Groups A, B, C, and D were given 0, 20, 60, and 200 preliminary trials, respectively.

METHOD

Apparatus.—The darkened experimental room contained five booths. The *E* sat in the middle booth while the four *S*s sat in the remaining booths. Each booth contained a 1.5-in. lever and two reinforcing lights arranged on the right and left above the lever. The lever could be moved to the right or left and contained a spring which returned it to center when released.

The sequence of the buzzer signal and reinforcing light was controlled by an automatic programming device. The responses of *S*s and the reinforcing events were recorded automatically.

Design.—The four experimental groups, 48 *S*s in each, were given 0, 20, 60, and 200 trials, respectively, under a 50:50 reinforcement schedule before being shifted to a 90:10 schedule. Reinforcement sequences were constructed randomly in blocks of 20 trials. For each group there were 12 different sequences, four *S*s to a sequence, and each sequence was matched as much as possible across groups. The side of the more frequent reinforcing light was counterbalanced within each subgroup of four *S*s. The order in which the 48 subgroups were run was determined randomly.

Subjects.—The 192 *S*s were students from elementary psychology courses at Indiana University.

Procedure.—When *S*s were seated in the booths, *E* turned off the overhead light and instructed them to predict which of the two lights would follow a buzzer on each trial. The instruction procedure followed closely that used by Estes and Straughan (1954), with the notable exception that practice trials were omitted in the present study.

Following the instructions, the sequence of 300 trials was given without interruption. The time relationships within trials were as follows: duration of the signal, 2 sec.; time between the signal and the reinforcing light, 1.2 sec.; duration of the reinforcing light, 1 sec.; time between the reinforcing light and the next signal, .4 sec.

RESULTS AND DISCUSSION

The learning curves under the 90:10 reinforcement condition for the four groups are shown in Fig. 2. The four curves appear to be ordered as predicted despite the reversal of the initial points of Curves B and C. The high initial level of Group C may be accounted for by chance factors, since this group was responding consistently above .50 toward the end of the preliminary trials (cf. Fig. 3, Group C). The measure of learning rate for each *S* is the mean number of A_1 predictions over the first 100 trials, which was 81.73, 77.75, 75.58, and 74.39 for Group A, B, C, and D, respectively. Analysis of variance of these group means yields an *F* of 3.70, with 3 and 144 *df*, which is significant at the 5% level (the σ_m with sequence error removed is 1.95).

The apparent similarity of rates for Groups C and D indicates that the effect of preliminary trials is approaching an asymptote. This result is expected on the theoretical grounds that the proportion of neutral elements decreases toward zero as trials increase.

Turning to the asymptotic features of the curves in Fig. 2, it may be noted that response proportions for Groups A, B, and C appear to converge to a common level. The ob-

tained proportions of A_1 responses over the last 40 trials for Groups A, B, and C are .901, .897, and .901, respectively. None of these values is significantly different from the predicted .90 by t tests. Similar determination for Group D was not feasible because it had not yet reached asymptote.

Details of the estimation procedure and theoretical curves for the present study may be found in an article concerned with a formal presentation of the model (LaBerge, in press).

Had the curves of Fig. 2 been plotted in terms of A_2 responses the resulting family of descending curves would resemble a set of extinction curves. Interpreted in this way, the results imply that the rate of extinction decreases as the number of reinforced trials under a partial reinforcement schedule increases. Although no parallel investigation of this effect in the free responding situation has been reported to date, experiments using periodic (Wilson, 1954) and fixed-ratio (Denny, Wells, & Maatsch, 1957) reinforcement schedules indicate that varying the number of trials under these schedules produces a similar effect upon extinction.

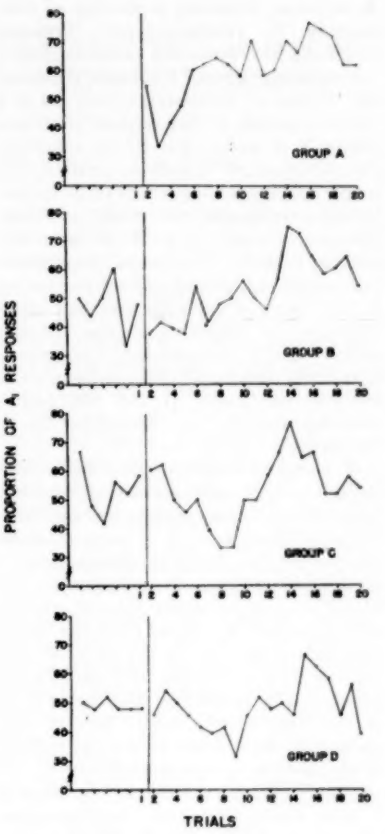

FIG. 3. Trial-by-trial plot of the proportion of A_1 responses for the four groups over the last five trials of the 50:50 condition and the first 20 trials of the 90:10 condition. Trials are numbered beginning with the first response of the 90:10 condition.

A persistent problem in fitting theoretical curves to data from prediction experiments arises within the first few learning trials. Typically the proportion of A_1 responses begins at or near .50, but instead of moving immediately upward toward asymptote, as predicted, it remains at .50 or even drops below for a few trials (Bush & Mosteller, 1955). When the present data were plotted trial by trial, as shown in Fig. 3, the "dip" effect at the beginning of the 90:10 condition may be observed for each group. It appears that the number of trials under a 50:50 reinforcement condition does not affect the depth of the "dip" but simply postpones its onset, and perhaps prolongs the effect. In its present form, the neutral elements model apparently cannot account for the "dip" effect any better than other models (Burke & Estes, 1957; Bush & Mosteller, 1955; Restle, 1957). However, we shall attempt in an *ad hoc* manner to explain the obtained occurrence and delay of the "dip" by assuming that S comes to the experiment with an alternation tendency, that is, a high probability of making an

A_2 response following a number of consecutive E_1 reinforcements. Training under the 50:50 schedule presents chance occurrences of runs of E_1 events, the number of runs of increasing length being a direct function of the number of trials. While *S* is under the 50:50 schedule, the alternation response emitted by longer and longer sequences of E_1 events should extinguish as trials increase. Therefore, when the 90:10 schedule with its long E_1 runs occurs, the alternation responses should appear only after E_1 sequences of a length that had small likelihood of occurring in the training series. And thus increasing the amount of training under a 50:50 schedule should increase the delay of the alternation response and the corresponding dip in the learning curve.

Although the delay of the "dip" tends to increase the differences in learning rates of the four groups in the first block, it is clear that it cannot alone account for the obtained differences over the first five blocks shown in Fig. 2. In Fig. 3 it may be noted that all groups appear to have recovered from the "dip" well before the end of the first 20-trial block. Even ignoring the differences between the groups in the latter half of the first block, comparison of the groups in Fig. 2 on the next several blocks shows that Groups B and C do not reach the level attained by Group A until Block 9 or 10. Furthermore, comparison of Groups B and C shows that even though Group C appears superior to Group B in Block 1, Group C drops below Group B for the next four blocks. Finally, the percentage of the distance to asymptote obtained in the second block by each group is 76%, 68%, 66%, and 56% for Groups A, B, C, and D, respectively. Therefore, there appears to be a consistently ordered difference in rate of learning by the groups beyond the first block of trials, as expected from the neutral elements model. Further research is indicated in order to separate the neutral elements effect from the presumed alternation effect in the first few trials of the prediction learning situation.

Summary

An analysis of rate of conditioning in a simple prediction situation was carried out to test a new assumption in statistical learning theory. From the assumption that some stimulus elements are initially neutral relative to all of the experimentally defined responses in a learning situation, it was predicted that rate of conditioning should decrease as the amount of preliminary training increases.

Instructions were to predict which of two lights would flash following a buzzer. Four groups of 48 *S*s received 0, 20, 60, or 200 trials under 50:50 reinforcement condition. Following these preliminary trials, all groups were shifted to a 90:10 condition for a series of trials. Resulting learning rates under the 90:10 condition for the four groups differed significantly and were ordered as predicted. Obtained and predicted asymptotes showed very close agreement.

Although a systematic delay in the "dip" effect was found to affect learning rate in the first few trials, it is tentatively concluded that the obtained effect of preliminary training upon learning rates in large part is accounted for by the neutral elements model.

References

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

DENNY, M. R., WELLS, R. H., & MAATSCH, J. L. Resistance to extinction as a function of the discrimination habit established during fixed-ratio reinforcement. *J. exp. Psychol.*, 1957, **54**, 451–456.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

LABERGE, D. L. A model with neutral elements. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, Calif.: Stanford Univer. Press, in press.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

WILSON, M. P. Periodic reinforcement interval and number of periodic reinforcements as parameters of response strength. *J. comp. physiol. Psychol.*, 1954, **47**, 51–56.

(Received March 3, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

RELATION OF STRESS AND DIFFERENTIAL POSITION HABITS TO PERFORMANCE IN MOTOR LEARNING

ALFRED CASTANEDA AND LEWIS P. LIPSITT[1]

Iowa Child Welfare Research Station, State University of Iowa

In several previous studies (Castaneda, 1956; Castaneda & Palermo, 1955; Palermo, 1957) stress (time pressures) has been introduced during a transfer task involving re-pairing of several S-R associations acquired during original learning while the remaining S-R associations are left unchanged. Under these conditions, the effects of stress have been found to depend on whether re-pairing is involved, i.e., stress has been found to impair performance where re-pairing is involved but to facilitate where it is not. On the assumption that this form of stress affects the level of drive in young human *S*s, these results have been interpreted to be consistent with those theoretical implications (Spence, 1956; Taylor, 1956) derived from Hull's (1943) assumption that all existing habits aroused by the stimulus situation are indiscriminately strengthened by the total drive level that is momentarily operative. Thus, if the stimulus situation gives rise to more than one response, the effects of increased drive, according to this assumption, are expected to depend on the strength of the to-be-learned (correct) response relative to other competing (incorrect) responses.

The present study is concerned with the relation of this form of stress to performance in a motor-learning task in which initially dominant (pre-experimentally acquired) position habits are arbitrarily designated either correct or incorrect. A schematized version of the task consisting of two horizontally parallel rows of eight lights and eight switches is presented in Fig. 1. The to-be-learned S-R (light-switch) combinations are indicated by means of the connecting arrows. The present version is a modification of one described by Morin and Grant (1955) which, in turn, is based on suggestions by Fitts and Seegar (1953). Morin and Grant have reported findings in line with the assumption that under conditions in which the various lights are successively activated the initially dominant response is to the response element which is in direct spatial correspondence, i.e., directly underneath. As can be seen in Fig. 1, half of the to-be-learned S-R combinations in the present situation (No. 1, 4, 5, and 8) involve the initially dominant response while the remaining half (No. 2, 3, 6, and 7) do not. According to the previously outlined assumptions, the effects of stress would also be expected to interact with the particular S-R combination involved. For those combinations in which the initially dominant response is compatible with the correct one, performance under stress should be facilitated in comparison to non-stress, but impaired where it is incompatible with the correct response.

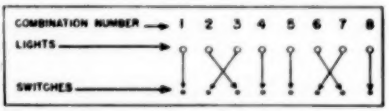

FIG. 1. Schema of the to-be-learned S-R (light-switch) combinations.

[1] Now at Brown University.

METHOD

Subjects.—The *Ss* were 108 fifth-grade boys and girls randomly assigned in equal numbers, 54 per group, to the stress and nonstress conditions.[1] The last 38 *Ss*, 19 per group, were added at a later time in order to obtain information as to the specific types of errors occurring to those S-R combinations (No. 2, 3, 6, and 7) in which the dominant position tendency was incompatible with the correct response. Records of these errors had not been obtained for the first 70 *Ss*.

Apparatus and procedure.—The apparatus consisted of a sloping panel 20 × 12 in., painted flat black. Eight green, jeweled reflectors, illuminated by 6.3-v. pilot lamps, were arranged horizontally on the panel spaced 2 in. apart. A small push button was situated 2 in. directly below each light. A 6.3-v. red, jeweled reflector was centered at the top of the panel. By means of appropriate switching devices on the back of the panel not visible to *S*, any single light could be activated and any single button could be set to turn it off. Selection of the correct button by *S* simultaneously terminated the green light and activated the red light which indicated that the response was correct. For the stress condition, duration of the stimulus light was electronically controlled and responses occurring after the automatic offset of the light did not activate the red signal lamp.

The *Ss* in the nonstress condition were informed that the task required them to learn which of the buttons turned off a given light and that for "some" of the lights the button directly underneath was correct but that for "some" of the lights an immediately *adjacent* button was correct. The intent in this portion of the instructions was to limit the number of response elements from which *S* was to make his selection. No time limits were placed on responding for the nonstress group. In addition to these instructions, *Ss* in the stress condition were given a demonstration in which a stimulus light was activated for 1 sec. and were informed that failure to respond within this interval would be scored an error. As in the previous studies, in order to reduce the incidence of nonresponding, the timer was set for a 2-sec. exposure during the test trials. If a response failed to occur during this interval, the light was re-presented until a response could be recorded for that trial. The data reported, however, do not include failures to respond. Two *Ss* each failed three times to respond on the first presentation of a light while seven *Ss* failed once. Three of these failures occurred on lights which were associated with the button directly underneath and 10 on lights involving an adjacent button.

Each light was presented 10 times. Defining each presentation for which a response was recorded as a trial, training consisted of a total of 80 trials. The lights were successively activated in the same prearranged order for all *Ss*. This consisted of one presentation of each light in every block of eight trials and no light was activated twice in succession (a qualification to this aspect of the procedure may be noted in the preceding paragraph). No attempt was made to hold the interval between presentations constant and *Ss* were *not* permitted to correct for errors.

Since principal interest is in the differential effects of stress in relation to the two types of S-R combinations, those combinations for which the correct response involves the button directly underneath will be hereafter designated DTC (dominant tendency correct), and those involving adjacent buttons designated DTI (dominant tendency incorrect).

RESULTS

Table 1 presents the percentage of correct responses made by the stress and nonstress groups on all trials to the DTC and DTI combinations separately. Figure 2 presents these data in blocks of 20 trials consisting of 10 trials on the DTC combinations and 10 on DTI. The ordinate values represent percentage of correct responses based on 10 trials. On the

TABLE 1

PERCENTAGE OF CORRECT RESPONSES

Group	S-R Combination			
	DTC		DTI	
	Mean	*SD*	Mean	*SD*
Stress	70.60	13.42	33.29	15.13
Nonstress	62.92	11.00	42.18	14.07

[1] Appreciation is expressed to H. C. Reid, Director of Curriculum and Instruction, Cedar Rapids Public Schools, for his generous assistance and cooperation.

basis of these data it is apparent that both groups made fewer correct responses on the DTI than the DTC combinations with performance starting out well below 50% on the DTI at the first block of trials. More importantly, however, these data indicate that the effects of stress are contingent on the type of S-R combination. Thus, it is apparent that the stress group, in comparison to the nonstress group, made more correct responses on the DTC combinations at all stages of training and, with the exception of the last block, fewer correct responses on the DTI. It is of interest to note that the point at which the DTI performance curves converge approximates the 50% level of performance. The mean percentage of correct responses on the last block of trials was 48.70 for the nonstress and 48.14 for the stress group.

A summary of an analysis of variance (Lindquist, 1953, pp. 291–297) of the data presented in Fig. 2 appears in Table 2. The differential effects of stress (facilitation, impairment) indicated in Table 1 and Fig. 2 is seen to be statistically reliable as reflected in the interaction of stress with the type of S-R combination.

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RESPONSES

Source	*df*	*MS*	*F*
Stress (C)	1	.78	
Error (b)	106	804.66	
Blocks (A)	3	167.13	84.40**
S-R Type (B)	1	1820.04	255.26**
A × B	3	7.38	2.64*
A × C	3	4.01	2.02
B × C	1	148.34	20.80**
A × B × C	3	6.48	2.32
$Error_1$ (w)	318	1.98	
$Error_2$ (w)	106	7.13	
$Error_3$ (w)	318	2.79	

* $P = .05$.
** $P = .001$.

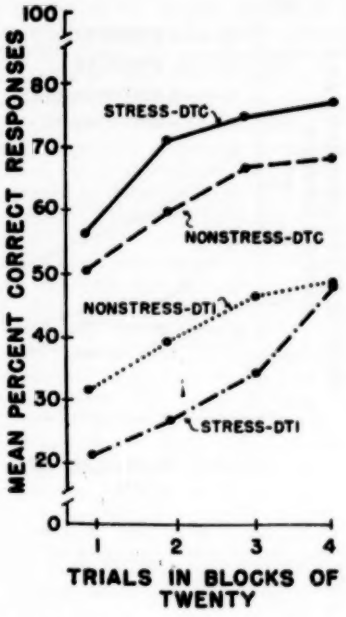

FIG. 2. Mean percentage of correct responses for each type of S-R combination plotted separately for the stress and nonstress groups.

Comparisons of the difference between the stress and nonstress groups on the two types of combinations separately, based on the means shown in Table 1, indicated that the difference on the DTC combinations was reliable at beyond the .005 level ($F = 10.64$, $df = 1$ and 106), and on the DTI at beyond the .005 level ($F = 9.77$). The apparent tendency for the effects of stress to depend on the stage of training on the DTI combinations would be reflected in the triple interaction. It can be noted, however, that the reliability of this effect falls just short of the .05 level. However, because of its theoretical implications and the possibility that this effect may have

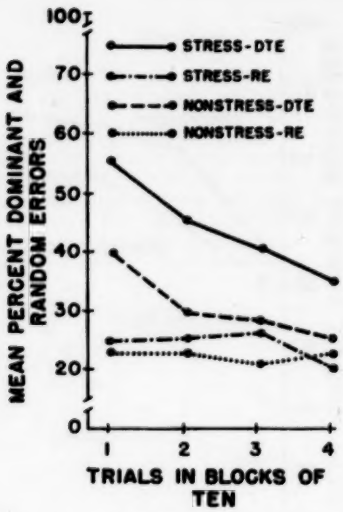

FIG. 3. Mean percentage of dominant tendency and random type errors occurring to the DTI S-R combinations plotted separately for the stress and nonstress groups.

been partially obscured in the over-all analysis, an additional analysis based on the two curves alone was performed with the main interest being in the interaction of the effects of stress with that of stage of training. This effect was found to be reliable at beyond the .05 level ($F = 3.20$, $df = 3$ and 318), providing evidence for rejection of the hypothesis of parallel curves.

While the finding indicating that the stress group made more correct responses than the nonstress group on the DTC combinations provides evidence that stress enhanced the strength of the dominant tendency, the opposite results on the DTI stimuli do not necessarily permit this same inference. That is to say, it is possible that this latter effect may have been the result of an increased frequency in responses to incorrect buttons other than the one in direct spatial correspondence. The data of the last 38 *Ss*, for whom records were kept of the type of error occurring to the DTI combinations, provide the relevant information. Responses occurring to the button in direct spatial correspondence are designated DTE (dominant tendency errors), and responses made to any other incorrect button designated RE (random errors).

Figure 3 presents the percentage occurrence of the two types of errors made by each group separately in blocks of 10 trials. On the basis of these data it is apparent that both groups made more errors of the DTE than the RE type with this difference being more pronounced during the early stages of training. It is also apparent that the stress group made consistently more errors of the DTE type than the nonstress group. On the other hand, the stress group also tended, but not consistently, to give slightly more errors of the RE type than the nonstress group. An analysis of variance of these data yielded three statistically reliable effects. These were Stress ($P < .025$), Blocks of Trials ($P < .001$) and the Type of Error ($P < .001$). Although, as can be seen in Fig. 3, the magnitude of the difference between the two groups appears to depend on the type of error, the interaction of stress and type of error did not reach reliability at the .05 level. However, when the difference between the two groups was compared on the total number of each type of error separately, the difference was reliable at beyond the .05 level for errors of the DTE type but considerably short of the .05 level on errors of the RE type.

Discussion

From a theoretical standpoint the results of the present experiment indicating differential effects of stress (facilitation, impairment) are of special interest and are in line with the previous studies. The effects of stress were found to be contingent on the relative strength of the to-be-learned response. For those S-R combinations in which the initially dominant response was arbitrarily designated correct, performance was found to be significantly facilitated under the stress as compared to the nonstress condition. On the other hand, deleterious consequences of stress were found for those combinations in which the initially dominant response was incompatible with that required for correct performance. In the latter case the results indicated that this effect could be partly attributed to the relatively greater enhancement in the strength of the initially dominant response.

Of additional interest for its possible theoretical implications was the finding that the magnitude of the deleterious effects of stress was dependent on the stage of training for those S-R combinations in which the initially dominant response was incompatible with that required for correct performance. As training progressed, the initial relative superiority of the nonstress condition tended to diminish so that at the end of training the performance of both groups was nearly identical. The fact that this equality in performance occurred when the frequency of correct responding approximated 50% is of theoretical interest. That is, assuming that such a level of performance indicates equality in the strengths of the correct and the initially dominant, incorrect tendency, i.e., $H+ = H-$, an increase in drive, according to theory, would not be expected to affect performance under these conditions, and neglecting other factors. Since drive is assumed to combine indiscriminately with *all* existing habit tendencies, the resultant effect would be one of augmenting the absolute strengths of the correct as well as the incorrect response by equal amounts. Providing that the equality in the habit strengths of the correct and incorrect responses occurs at approximately the same stage in training for both a high- and a low-drive condition, performance under the higher drive would not be expected to differ from that of low-drive condition. However, once the strength of the correct habit, though initially the weaker, supersedes the strength of the incorrect one through the process of differential reinforcement, performance under the higher drive condition would be expected to be facilitated thereafter. Thus, on the basis of these implications, had training been continued in the present experiment, eventual facilitation under stress would be expected on those S-R combinations in which the initially dominant tendency was incompatible with the correct response even though its initial effect is that of impairment. As implied by the theory, the change from relative inferiority to superiority in performance would be expected to occur beyond that point in training when the strengths of the correct and incorrect responses have become equal.

Summary

The present experiment was concerned with the relation of time pressures (stress) and pre-experimentally acquired differential position habits to performance of children in a motor-learning task. The main interest was in the effects of stress as a function of the compatibility of the dominant position habits with the correct response in the task. An interaction between stress and compatibility was found based on statistically reliable effects for stress to facilitate performance, in comparison to a nonstress condition, where the dominant position habit was compatible with the correct response but to interfere where it was incompatible. The experiment was designed to test implications relating to the role of drive in situations involving competing responses based on Hull's basic assumption concerning the manner in which drive and habit combine to determine response strength.

References

CASTANEDA, A. Effects of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.

CASTANEDA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

FITTS, P. M., & SEEGAR, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MORIN, R. E., & GRANT, D. A. Learning and performance on a key-pressing task as a function of the degree of spatial stimulus-response correspondence. *J. exp. Psychol.*, 1955, **49**, 39–47.

PALERMO, D. S. Proactive interference and facilitation as a function of amount of training and stress. *J. exp. Psychol.*, 1957, **53**, 293–296.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

(Received February 14, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

THE DIFFERENTIAL EFFECTS OF WORD AND OBJECT STIMULI ON THE LEARNING OF PAIRED ASSOCIATES[1]

C. C. WIMER[2] AND W. E. LAMBERT

McGill University

The present study is concerned with a problem which has arisen from the contention of language-learning theorists that foreign words are learned more efficiently when associated directly with environmental events than when paired with their equivalents in the native language. If this is true, analogous results should be obtained when nonsense syllables are learned by the paired-associates method in a controlled experiment. Nonsense syllables, as responses, should be learned faster when paired with object stimuli than when the stimuli are names of objects.

Early experiments which attempted to confirm the hypothesis yielded conflicting results. Most of these experiments were carried out in an actual classroom language-learning situation, with secondary school pupils and college students as *Ss*. Characteristically, the number of *Ss* was small, presentation was uncontrolled, and statistical analysis was inadequate. Three experiments reported by Huse (1931) obtained quite different sets of results. One, a study by Peterson in 1903, found that recall of nonsense responses was better when they had been paired with object stimuli than when the stimuli were nouns. A second, by Schlueter in 1914, corroborated the above findings but obtained better *learning* scores for nonsense responses paired with word stimuli. A third study (Netschajeff, 1908), the only one in which the experimental conditions were rigorously controlled, found that nonsense responses were both learned faster and retained better over long periods of time when paired with word stimuli than when the stimulus member was an object. A more recent study (Kopstein & Roshal, 1954) showed that a foreign vocabulary was remembered better when learned from pictures than when associated with words, giving indirect support to the hypothesis.

In view of the ambiguity of the above results, the present experiments were undertaken in the hope that a thorough investigation of the differential effects on verbal learning of word and object stimuli might throw some light on the problem, as well as yielding additional insights into the processes underlying the learning of verbal materials.

Several theories have been proposed to account for the relative difficulty for learning of two lists of paired associates. Two hypotheses in particular seem applicable to the differences between word and object stimuli.

First, it is generally agreed that an increase in the meaningfulness of a stimulus increases the speed of learning of a response paired with it. Noble (1952a, 1952b) and Noble and McNeely (1957) have defined "meaningfulness" as the average number of written associations given to a verbal stimulus during a 60-sec. time interval. His experiments show that difficulty in verbal learning decreases as m, the average meaningfulness of a list of stimuli, increases.

Second, according to Gibson (1940, 1942) the difficulty of learning a list of paired associates is proportional to the extent of intralist generalization. Generalization is defined as the ten-

[1] This study, a part of a program of research on bilingualism, was supported by the Canadian Defence Research Board, Grant No. D77-94-01-10. The writers wish to thank Eleanor J. Gibson for her suggestions concerning the preparation of this paper.

[2] Now at Rutgers University.

dency for a response, R_a, originally paired with a stimulus, S_a, to occur as a response to S_b, where S_a and S_b are similar in some respect.

The present study consists of three experiments. Experiment I is an attempt to determine whether nonsense-syllable responses are learned faster when paired with word or with object stimuli; Exp. II is an analysis of the differences in meaningfulness between objects and their names; and Exp. III is a comparison of the intralist similarity of a set of objects and a set of words. The method used to measure similarity is derived from the semantic differential (Osgood, Suci, & Tannenbaum, 1957), and is based on the assumption that the average semantic distance separating the stimuli in a list is a good measure of intralist similarity, a small distance representing a large amount of similarity, and vice versa.

Method

Subjects.—The *Ss* in the three experiments were undergraduate psychology students at McGill University. Sixty *Ss* took part in Exp. I; 20 learned List W (described below), 20 learned List O, and 20 learned List M. For Exp. II, there were 24 *Ss*; 12 gave associations for the stimuli of List W, and 12 associated to List O. In Exp. III, 36 *Ss* gave semantic differential ratings for the stimuli; 12 rated the stimuli of List W, 12 rated List O, and 12 rated List M. No *S* took part in more than one experiment. The three experiments were run successively, and no attempt was made to equate *Ss* throughout. The abilities required by the tasks were so different that no practice test could be found which was relevant to all three. Since all *Ss* were drawn from a fairly homogeneous population, and since no direct comparison between experiments was intended, it was decided that between-experiments equating of *Ss* was unnecessary.

Materials.—For Exp. I, three lists of nine paired associates each were compiled. For List O, the stimulus members were nine common small objects; for List W, the stimuli were the names of the nine objects; and for List M, there were nine mixed stimuli, i.e., some from List W and some from List O. The response members were the same for all three lists—nine nonsense syllables of less than 35% association value (Hilgard, 1951). Each syllable was paired with equivalent stimuli in all lists. A list of the stimuli used and the nonsense syllables with which they were paired is presented in Table 1. The stimuli were typed or mounted on 3 × 5-in. white cards; the response syllables were typed on the right. Half the cards contained only the stimulus members of the pairs, and half had both stimulus and response members. The cards were then divided into three sets of 54 each, each set consisting of three different orders of a list. The cards were mounted in a Kardex Book Unit in such a way that they could be turned over one by one at a constant rate.

For Exp. II and III, the same materials were used, except that the cards shown were those containing only the stimulus members of the pairs. For the semantic differential used in Exp. III, Form II, as suggested by Osgood, et al. (1957, p. 81), was used. Of the 20 scales employed, 12 were chosen from the evaluation, potency, and activity categories, the three most

TABLE 1

List of Stimuli and Nonsense-Syllable Responses

Stimulus		Nonsense-Syllable Responses (Both Lists) and Association Values
List W	List O	
ring	gold-colored ring with green stone	vud (27%)
pencil	short red pencil	dax (0%)
match	burnt wooden match	bip (20%)
mirror	oval hand mirror	keb (7%)
star	5-pointed star outlined on index card	zot (7%)
button	large plain black button	maf (27%)
string	short piece of white string	nej (20%)
bandage	plastic "band aid"	yil (0%)
key	un-notched gold key	jom (33%)

important as determined by several factor analyses (Osgood et al., 1957). The criterion for the remaining eight scales was applicability for objects. Eight of the scales represented physically descriptive continua (e.g., large–small); the other 12 adjective pairs described more abstract dimensions (e.g., pleasurable–painful).

Procedure.—In Exp. I, all *Ss* were first given a practice list, consisting of four pairs, two word-syllable pairs and two object-syllable pairs. Each pair was presented four times. Subsequently, *Ss* were assigned alternately to List W, O, or M, which was learned to a criterion of one errorless trial. All responses were recorded by *E*. The interval between presentation of stimulus and response, and between a response and the following stimulus, was 3 sec., timed by a metronome. There was no additional intertrial interval, the last response of one order of a list being followed in 3 sec. by the first stimulus of the next.

For Exp. II, *Ss* were assigned randomly to either List W or List O, and were asked to give written associations to the stimuli. The stimuli were presented one by one, and each was exposed for 60 sec.

In Exp. III, *Ss* were assigned alternately to Lists W, O, and M, for which they gave semantic ratings. Each stimulus was shown individually and remained exposed until it had been rated on all 20 scales. All *Ss* received 10 stimuli altogether, the first of which was considered a practice trial. This first trial involved rating a concept which was not part of the original list and was thus not taken into account in the statistical analysis.

Results

Before analysis of Exp. I was undertaken, performance on the practice task was analyzed in order to determine whether differences between List W and List O could be accounted for by group differences in learning ability. The mean number of errors for Group W was 4.95 ($SD = 2.92$), and for Group O, 5.00 ($SD = 3.03$); $t < 1$. The correlation between errors on the practice list and errors in acquisition of the main list was .39, $P < .02$ for $n = 40$. It seems reasonable to conclude that the two groups were equal with respect to learning ability and that any differences obtained can be attributed to the experimental treatment.

Table 2 shows the means for Lists W and O for each experiment. It can be seen that in Exp. I fewer trials were required to learn List O, and that there were fewer errors (failures to respond correctly) in acquisition of List O.

According to the hypotheses discussed in the introduction, one would expect the direction of these results to be reflected in the other two experiments, as follows: If the greater difficulty of List W is due to the fact that objects are more meaningful than their names, the mean number of associations for List O should be significantly greater than the mean for List W. If the differential effects are due to less similarity of meaning among objects than among their names, the mean D_{ij} (average distance separating stimuli) will be significantly greater for List O than for List W.

For the association task of Exp. II, it can be seen in Table 2 that the direction is not as predicted nor is the difference significant. Thus the differences in learning cannot be accounted for by the greater meaningfulness of object stimuli.

In the analysis of Exp. III, a method suggested by Osgood, et al. (1957, p. 102) was used to compare the average distances separating stimuli: $D_{ij} = \sqrt{\sum d_{ij}^2}$, where D_{ij} represents the distance between concepts i and j, and d_{ij} represents the algebraic difference between the mean group ratings for i and j on any given dimension (Osgood et al., p. 91). D_{ij} was computed for all possible pairs of concepts $\frac{n(n-1)}{2}$ pairs, where n = number of concepts) for List W, and for List O. Subsequently D_{ij}'s were determined separately for the 8 physical scales and for the 12 abstract scales, and the two lists were compared for both sets of D_{ij}'s. Since

TABLE 2

MEAN SCORES ON EACH TASK

Exp.	Variable	List W		List O		Test	df	P
		Mean	SD	Mean	SD			
I	Trials to criterion	8.35	3.77	6.05	2.22	$t^* = 2.29$	19	<.05
	Errors during learning	40.95	23.73	25.30	11.00	$t^* = 2.61$	19	<.02
II	Number of associations per stimulus	9.71	2.21	8.31	2.25	$t = 1.47$	22	n.s.
III	D_{ij} All scales	5.28		6.32		$\chi^2 = 6.92$	1	<.01
	D_{ij} Physical scales	4.45		5.42		$\chi^2 = 5.56$	1	<.02
	D_{ij} Abstract scales	2.68		3.13		$\chi^2 = 0.88$	1	n.s.

* t corrected for heterogeneity of variance (Edwards, 1950, p. 170). Since there were the same number of *Ss* in both groups, the correction consisted of halving the df.

the distribution of D is unknown, the median test (Mosteller & Bush, 1954, p. 314) was used as an estimate of the significance of the difference between lists. The analysis supports the hypothesis that there is less intralist similarity among the stimuli of List O, as shown in Table 2. The separate analysis of physical and abstract scales shows that most of the differences between the meanings of objects and their names are differences in the physical aspects of the concepts aroused.

Table 3 shows the mean scores for List M, the mixed list, in which there were significantly more errors than in List O. The D_{ij} for List M, however, is significantly greater than that for List W, and about the same as that for List O. In this case, intralist semantic distance is not a good predictor of errors in learning. List M was included in the experiment in order to determine whether the differences between List W and List O would be reflected in a list containing both types of stimuli. (Within List M, there were significantly fewer errors for object-syllable pairs than for word-syllable pairs, $P < .001$.) It was not originally intended that a comparison should be made between this list and the other two, because of the differing amounts of relevant practice—the four practice pairs contained two word and two object stimuli, so that the practice list was, in effect, a miniature List M, whereas only two of the pairs were of the type that would occur in List W or List O. However, the fact that List M *does* appear to be more difficult than List O, while not showing a lower D_{ij}, limits the generality of the hypothesis that difficulty is reflected in semantic distance.

TABLE 3

MEAN SCORES FOR LIST M

Variable	Mean	Comparison With	
		List W	List O
Errors during learning	48.50	$t<1$, n.s.	$t^*=3.55$ $P<.01$
Mean D_{ij}	6.58	$\chi^2=13.50$ $P<.001$	$\chi^2=.67$, n.s.

* t corrected for heterogeneity of variance (Edwards, 1950, p. 170).

DISCUSSION

Two questions arising from the study will be considered: First, What are the implications of the results of the present

experiments for second-language learning? And second, What are the reasons for the differential effects of word and object stimuli on verbal learning?

The fact that, under the conditions used here, object-word pairs were learned faster than word-word pairs suggests that environmental events are more effective stimuli for the acquisition of foreign-word responses than are native-language equivalents for the new words, at least for the learning of a simple, basic vocabulary. However, the generality of these conclusions and the true validity of this approach to second-language teaching can only be determined by ascertaining whether the technique produces the desired results over a long period of time, and by testing its practical applicability in the classroom situation.

The explanation given here for the greater effectiveness of object stimuli—i.e., that there is less meaningful similarity among objects than among their names—is clearly not the only possible explanation. The hypothesis that faster learning of objects is due to their greater meaningfulness has been tested and rejected here, but several other theories remain to be considered. While generalization among meaningful verbal stimuli seems to be dependent on semantic factors (Osgood, 1953; Underwood, 1953b), formal or physical similarity has been found to affect the learning of comparatively meaningless materials (Gibson, 1942; Underwood, 1953a). The physical similarity among objects appears to be much less than that among object names, and it should not be disregarded as a possible alternative or supplementary explanation. There is also a possibility that generalization among stimuli and responses is greater for a list of word-syllable pairs, i.e., that the necessary discrimination between stimulus and response is more difficult when both are verbal items.

It might be said, then, that the lower semantic similarity among objects is a sufficient explanation for their greater effectiveness as stimuli, although it may not be the only adequate one. It is further restricted by the fact that the differences between words and objects are limited to purely physical meaning-dimensions. It should be noted that to say that two *concepts* are similar or different on a physical basis is not to say that two *stimuli* are physically similar or different. The difference is not in the stimuli themselves, but rather in the meanings attached to the stimuli. For example, consider three stimulus words, "pin," "pan," and "needle." Comparing the three words on a physical basis, with no reference to their meanings, one is forced to conclude that "pin" is more similar to "pan" than it is to "needle," with respect to size, component elements, etc. However, comparing the *meanings* of the three words, i.e., the concepts they arouse, one finds "pin" to be more similar to "needle" than to "pan," with respect to the *same* physical categories—size, component elements, etc. It is this second type of similarity that leads to generalization among materials of the type used in this experiment.

The use of the semantic differential to predict the difficulty of the particular materials used here suggests that it may be generally useful in the calibration of the difficulty of verbal lists.

Summary

The purpose of the present experiment was to compare the effects of word and object stimuli on the learning of nonsense-syllable responses. In Exp. I, three lists of paired associates, with either objects, words, or a combination of the two as stimuli, were learned by 60 college students. The responses for all lists were nonsense syllables. It was found that significantly fewer errors were made in the acquisition of the object-syllable list than in the word-syllable or the mixed list. In Exp. II, the meaningfulness (Noble's *m*) of each of the stimuli was tested by determining the average number of associations given by 24 *Ss* to a stimulus in a 60-sec. interval. Results showed no significant differences in the amount of meaningfulness of words and objects. In Exp. III, the intralist stimulus similarity of the lists was measured by having 36 *Ss* rate their concepts of the words and objects on the semantic differential. The intralist similarity of the word-syllable list was significantly greater than that of the object-syllable list.

The results were interpreted in relation to verbal learning theory and a new measure of generalization among verbal materials was proposed.

References

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Huse, H. R. *The psychology of foreign language study.* Chapel Hill: Univer. of North Carolina, 1931.

Kopstein, F. F., & Roshal, S. M. Learning foreign vocabulary from pictures vs. words. *Amer. Psychologist*, 1954, **9**, 407–408. (Abstract)

Mosteller, F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. *Theory and method.* Cambridge: Addison-Wesley, 1954.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430. (a)

Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446. (b)

Noble, C. E., & McNeely, D. A. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. of Illinois, 1957.

Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142. (a)

Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (b)

(Received February 17, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

DECISION-MAKING BEHAVIOR IN A TWO-CHOICE UNCERTAIN OUTCOME SITUATION [1]

SIDNEY SIEGEL AND DONALD AARON GOLDSTEIN

Pennsylvania State University

The predictions people make when placed in a two-choice uncertain outcome situation have received considerable attention in recent years. In the classical situation, as first used by Humphreys (1939), *S* is asked to predict which of two events (e.g., illumination or nonillumination of a light) will occur after a signal stimulus. The two events occur with fixed but unequal probabilities, say π_1 and π_2, in a random sequence for a number of trials, and the occurrence of either is not contingent on *S*'s behavior. The *S* is instructed to do his best to predict which will occur. He is allowed to witness the event, and thus to determine for himself the correctness or incorrectness of his prediction.

There are two theoretical models of interest here which yield predictions about *S*'s behavior in a situation of this sort.

One is Estes' (1950, 1954; Estes & Burke, 1953) statistical learning model which yields the prediction that *S*s will learn to match their response ratios (the relative frequencies with which they predict each of the events) to the actual probabilities of occurrence of the events. The same prediction is yielded by the Bush-Mosteller (1955) model when certain restrictions are applied to the parameters of that model. A number of studies report findings in support of this prediction (Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951; Hake & Hyman, 1953; Jarvik, 1951).

The second model, formalized by Von Neumann and Morgenstern (1947), is a game-theoretic model. According to some interpreters, one prediction consistent with this model is that a person will learn to maximize the expected frequency of correct predictions. This is accomplished by his predicting the more frequent event on all trials. Evidence giving some support to this prediction may be found elsewhere (Galanter & Gerstenhaber, 1956; Goodnow, 1955; Toda, 1956).

From decision-making theory (Davidson, Suppes, & Siegel, 1957; Edwards, 1954), an hypothesis of *maximization of expected utility* may be drawn which will account for both sorts of prediction behaviors (matching response ratios and maximizing the expected frequency of correct predictions). The "utility" of an outcome is its subjective value. According to the decision-making approach, *S*'s choice behavior (here, his predictions) depends on certain conditions related to the reinforcement inherent in the situation. The general hypothesis is that a person will maximize expected utility in any case. It must be pointed out, however, that the components of the total utility vary in magnitude from one reinforcement situation to another, and thus the strategy to maximize expected utility varies.

It is reasonable to suppose that when *S* is in a situation in which the

[1] This investigation was supported by a research grant (M-1328) from the National Institute of Mental Health, U. S. Public Health Service.

only payoff attached to the outcomes is the satisfaction of having his prediction confirmed by the event or the dissatisfaction of having his prediction dis-confirmed by it, making a correct prediction of the rarer event has greater utility for *S* than making a correct prediction of the more frequent event. The *S* may be supposed to derive satisfaction from "playing a game with the machine," trying to "outwit" it. Moreover, it may be monotonous to make the same prediction on trial after trial over a long series, and predicting the rarer event may therefore satisfy *S*'s utility of relieving boredom. If it is true that among the components of *S*'s total utility in this "no-payoff" situation are the utility of correctly predicting the more frequent event, the (possibly greater) utility of correctly predicting the rarer event, and the utility of relieving boredom, then *S* may maximize his total satisfaction in this situation by choosing a "mixed strategy," and this may approximate matching his response ratio to the actual probabilities of occurrence of the two events. If such an account is correct, then a decision-making model and the Estes model could yield the same predictions concerning behavior in a two-choice no-payoff situation.

If, however, the utility attached to correct predictions is increased by a change in the conditions of the game (i.e., if there is potential satisfaction beyond that of having an event confirm one's prediction), or if the cost attached to an incorrect prediction is increased, then the prediction yielded by the hypothesis of maximization of expected utility diverges from the prediction yielded by the Estes model. That is, if the decision-making approach proposed here is correct, the introduction of systematic variation in the utilities attached to correct or incorrect predictions should be followed by systematic differences in *S*'s responses. As the utility of a correct prediction is increased or the negative utility of an incorrect prediction is increased, *S*'s prediction of the more frequent event should tend to 100%—i.e., he should tend to choose a "pure strategy" rather than a mixed one.

In the experiment designed to test this contention, three different experimental conditions (levels of reinforcement) were selected to effect systematic variation in the utility attached to correct and incorrect predictions. Under the "no-payoff" condition, the correctness or incorrectness of *S*'s predictions did not affect his monetary holding. Under the "reward" condition, the utility attached to a correct prediction was greater, for there was a monetary payoff for each correct prediction. Under the "risk" condition, the utility attached to a correct prediction was still greater, for there was a monetary loss for each incorrect prediction as well as a monetary gain for each correct prediction, and thus a correct prediction rather than an incorrect prediction had twice as great an effect on *S*'s monetary holding under the "risk" condition as under the "reward" condition.

The purpose of the procedure to be described was to provide data to test the hypothesis that *the asymptotic probability of S's predicting the occurrence of the more frequent event in a two-choice uncertain outcome situation is a function of the level of reinforcement present in the situation, such that the probability of predicting the more frequent event will tend toward unity as the rewards (positive utility) and costs (negative utility) of correct and incorrect predictions are increased.*

That is, using Estes' notation (cf.

Estes, 1950), where $\bar{p}_1(\infty)$ is the mean asymptotic probability of predicting E_1 for a group of like *S*s, the hypothesis is that $\bar{p}_1(\infty)$ under risk $> \bar{p}_1(\infty)$ under reward $> \bar{p}_1(\infty)$ under no payoff.

Method

Subjects.—The *S*s were 36 male undergraduate students at Pennsylvania State University.

Apparatus.—Each *S* was run individually. For the experimental session, *S* was seated before a board to which three electric light bulbs were attached. Two, the "event" bulbs, were 60-w. bulbs set 8 in. apart. Directly between these and 6 in. below was a red 7½-w. bulb which served as the signal stimulus (warning light). On the table in front of *S*, between him and the board, were two push buttons set 4 in. apart.

The *E* was seated behind the board. Unseen by *S*, his push buttons were connected to bulbs at *E*'s desk. Depending on which button *S* pressed, one of the two bulbs at *E*'s desk would light, enabling *E* to record *S*'s predictions for the trial. As will be discussed, the sequence of "event" lights (right or left) for the successive trials was determined in advance of a session, and *E* followed this sequence in throwing a switch from his desk causing one or the other "event" bulb to light. None of *E*'s switching or recording equipment was visible to *S*.

For each trial, the signal stimulus (red bulb) would light and remain on for 3 sec. During this time, *S* made his prediction by pressing one of the buttons on the table before him. When the red bulb went off, one of the "event" bulbs would light and would remain on for 2 sec. A 2-sec. period intervened between trials.

Procedure.—The 36 *S*s were randomly assigned to three groups of equal size, the group assignment determining the condition under which each *S* would be run. The three conditions were No Payoff, Reward, and Risk.

Under the No Payoff condition, the reinforcement for each prediction consisted simply in seeing the outcome, i.e., seeing whether the right or left bulb lighted, and thereby determining whether one's prediction was confirmed or dis-confirmed. The No Payoff condition has been the "classical" situation for studies of human behavior in two-choice situations.

Under the Reward condition, the reinforcement for each prediction consisted in seeing the outcome and receiving 5 cents for each correct prediction. The 5-cent reward was given at the conclusion of each trial in which *S*'s prediction was confirmed.

Under the Risk condition, the reinforcement for each prediction consisted in seeing the outcome, receiving 5 cents if the prediction was confirmed, and losing 5 cents if it was dis-confirmed. At the conclusion of every trial, *S* either received or forfeited 5 cents, depending on whether his prediction had been correct or incorrect.

Two series of event sequences were prepared. One contained 75 lefts and 25 rights; the other was identical except that the lefts and rights were reversed, so that it contained 25 lefts and 75 rights. To counterbalance for the effects of possible right-left preference, half of the *S*s under each condition were presented with each series.

In the 100-trial series, the order of occurrence of the two events was randomized, by use of a table of random numbers, with two modifications: adjustments were made so that (*a*) there was no run of more than six successive occurrences of the more frequent event, and (*b*) within every 20-trial block the 3:1 proportion was maintained.

Each *S* was seen at an individual testing session in which a 100-trial series was presented. Each such session required about 30-min.

At the start of the session, standard instructions were given to all *S*s: "Your task in this experiment is to try to predict correctly which light will come on following the onset of the signal light. Three seconds after the signal light comes on, either the left or the right light will come on. As it is very important that you make a choice for each trial, please make your prediction as soon as possible after the signal light comes on. In any case, you must make your prediction in less than 3 sec.

"On the platform before you there are two buttons. If your prediction is that the left light will come on, press the left button. If your prediction is that the right light will come on, press the right button.

"Do you have any questions as to what you are to do? Let's run through what could happen. I will give you an example of each event."

In addition to these standard instructions and examples given to all *S*s, some additional special instructions were given depending on the condition under which the *S* was being run. The No Payoff *S*s were told simply: "Now, do your best to predict correctly which light will come on for every trial." The Reward *S*s received these additional instructions: "Now, every time you predict correctly you will win 5 cents; every time you predict incorrectly you will not win anything. At the end of the session all the money you have won will be yours to keep. That is, for every time you predict correctly you will win 5 cents, and all the money in your possession at the end of the session will belong

to you. Now, if you make no correct predictions you will have won no money by the end of the game, but if you play carefully you should come out winning at the end of the session."

The additional instructions given to the Risk *Ss* were these: "Now, every time you predict correctly you will win 5 cents; every time you predict incorrectly you will lose 5 cents. We will start you off with 75 cents, and you will keep all of your winnings at the end of the session. That is, at the end of the session all the money in your possession will belong to you. Now, if you do not play carefully, you are likely to lose your starting stake and therefore end the session with no money. However, if you play carefully you could end the session with a considerable amount of money which would be yours to keep."

The instructions to both the Reward and the Risk *Ss* concluded with this statement:

"Let me make it quite clear: all the money you win you can keep. There are no strings attached, and please understand you will be paid the exact amount you win. This money does not belong to me; it comes from a research grant, so I am not concerned with how much or how little you win. The amount you win will depend entirely on how you play. Is that clear? All right. Now do your best to predict correctly which light will come on for every trial."

Every *S* under every condition was given 75 cents at the start of each session.

Results and Discussion

For the test of the hypothesis, *Ss*' final 20 predictions in the course of each 100-trial series were observed. The score for any *S* was the proportion of times he predicted (whether correctly or not) the occurrence of the more frequent event during those final 20 trials.

Table 1 shows the distribution of these scores for all *Ss* for the final 20 trials of the 100-trial series. Inasmuch as the research hypothesis was an ordered alternative hypothesis, these data were tested by the Jonckheere (1954) test, a nonparametric *k*-sample test against ordered alternatives. By this test, it is possible to reject the null hypothesis of no differences in favor of the alternative hypothesis at the $p < .00003$ level.

TABLE 1

Number of *Ss* Predicting More Frequent Event at Various Proportions During Final 20 Trials of First 100-Trial Series

Proportion of Predictions of More Frequent Event	Condition		
	No Payoff ($N = 12$)	Reward ($N = 12$)	Risk ($N = 12$)
.60	2		
.65	3		
.70	3	3	
.75	2	4	
.80	2	3	
.85		1	3
.90		1	1
.95			5
1.00			3

That is, the data confirm the prediction that *Ss* under the Risk condition will predict the more frequent light oftener than *Ss* under the Reward condition, and these in turn will predict the more frequent light oftener than *Ss* under the No Payoff condition. An interesting feature of the distributions shown in Table 1 is that there is no overlap between the scores of the Risk and No Payoff *Ss*.

The mean proportion of times that the more frequent light was predicted during the final trials of the first 100-trial series under the No Payoff condition was .70. Under Reward, the mean was .77. Under Risk, it was .93.

Four *Ss* were selected at random from the 12 under each condition and were run under the same condition for an additional 200 trials at a second session a week later. Although the number of cases is small, it is of interest to observe that these *Ss*' scores, during the final 20 trials of the second and third 100-trial series separately, each confirm the hypothesis, by the Jonckheere test, at well beyond the .01 level. For descriptive purposes, the mean proportions for

these subgroups are presented in Table 2. The body of Table 2 gives the average proportion of trials on which the more frequent light was predicted by these *Ss* in the various series under the various conditions. It is of some interest to observe that under both the No Payoff and the Reward conditions, these proportions increase from series to series (reaching $\pi_1 = .75$ for the No Payoff condition at the end of the third series), whereas under the Risk condition *Ss* exhibited stable behavior, on the average, throughout the three series. It would seem that *Ss* learn most rapidly under a condition of Risk (increased reinforcement).

The utility hinging on correct or incorrect predictions differed under the No Payoff, Reward, and Risk conditions, and thus the attempt to maximize expected utility led to different strategies under these conditions. Thus any attempt to construct a unified theory to account for the choice behavior to which both the Estes model and the game theoretic model have reference would have to consider, in addition to the utility of a correct prediction, other components of the total utility such as the specific utility of gambling, the negative utility of "boredom" (i.e., the result of always pressing the same button in predicting the same event over hundreds of trials), etc. If these components of the total utility are conceived as varying in magnitude under the experimental conditions, then the data which have been reported support the general hypothesis that *Ss*' choice behavior reflects an attempt to maximize expected utility.

Some investigators in this field, especially those influenced by game theoretic principles, have asserted that people who match their response ratios to the probabilities are acting "irrationally," because such people could increase their expected proportion of correct predictions by predicting the more frequent event on every trial. However, as Simon (1956, p. 271) has reminded us, one must bear in mind the distinction between *subjective* rationality (i.e., behavior that is rational, given the perceptual and evaluational premises of *S*) and *objective* rationality (rationality as viewed by *E*). The empirical fact that *Ss*' proportions of choices of the more frequent event increase as reinforcement and risk are increased seems to show that, in their choice behavior, *Ss* act as though they were attempting to maximize expected utility. If this is the case we can designate the mixed strategy behavior, in the No Payoff condition, as rational in that it will lead to the greatest expected payoff in total satisfaction.

TABLE 2

MEAN PROPORTION OF TIMES THE MORE FREQUENT EVENT WAS PREDICTED BY A SUBGROUP OF FOUR *Ss*, RANDOMLY SELECTED FROM EACH PAYOFF GROUP, DURING FINAL 20 TRIALS OF EACH 100-TRIAL SERIES

Series	No Payoff	Reward	Risk
1–100	.69	.78	.95
101–200	.74	.85	.95
201–300	.75	.86	.95

SUMMARY

The purpose of the experiment which has been reported was to test the hypothesis that the asymptotic probability of *S*'s predicting the occurrence of the more frequent event in a two-choice uncertain outcome situation is a function of the level of reinforcement present in the situation, such that the probability of predicting the more frequent event will tend toward unity as the rewards (positive utility) and costs (negative utility) of correct and incorrect predictions are increased. This hypothesis is drawn from the theory of decision making.

The *Ss*' predictions were observed in a two-choice uncertain outcome situation under three conditions: No Payoff, Reward, and Risk. The proportion of times *Ss* predicted the more frequent outcome differed under these three conditions in the predicted direction.

References

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Davidson, D., Suppes, P., & Siegel, S. *Decision-making: An experimental approach.* Stanford, Calif.: Stanford Univer. Press, 1957.

Edwards, W. The theory of decision-making. *Psychol. Bull.*, 1954, **51**, 380–417.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K. Individual behavior in uncertain outcome situations: An interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954. Pp. 127–137.

Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Galanter, E., & Gerstenhaber, M. On thought: The extrinsic theory. *Psychol. Rev.*, 1956, **63**, 218–228.

Goodnow, J. J. Determinants of choice distributions in two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

Jonckheere, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

Simon, H. A. A comparison of game theory and learning theory. *Psychometrika*, 1956, **21**, 267–272.

Toda, M. Guessing sequences under various conditions of payoff. *Jap. psychol. Research*, 1956, **4**, 11–23.

Von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior.* (2nd Ed.) Princeton, N. J.: Princeton Univer. Press, 1947.

(Received February 17, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

EXTINCTION AFTER ACQUISITION UNDER DIFFERENT VERBAL REINFORCEMENT COMBINATIONS [1]

ALEXANDER M. BUCHWALD

Indiana University

In the typical human learning experiment *S* is given instructions which specifiy how *E* will indicate whether a response is "correct" or "incorrect." For example, *S* might be told that a light flashing after a response indicates a "correct" response while failure of the light to flash indicates an "incorrect" response. If these instructions were omitted, and the flashing of the light were made contingent upon the occurrence of a particular response, the probability of that response might remain constant over a series of trials for some *S*s, while for others it might approach a value of one, or a value of zero. In these cases the data could be reasonably interpreted as indicating that the flash had no reinforcement value (was neutral), had a positive reinforcement value, or a negative reinforcement value, respectively. If, instead of presence or absence of a flashing light, the events which are contingent upon *S*s responses are verbalizations by *E*, comparisons of learning data for *S*s learning under different verbal reinforcers can indicate differences in the reinforcement values of these verbalizations for *S* populations. However, any such experiment requires a minimum of two reinforcing events and this simple comparison yields data concerning the relative reinforcing value of the reinforcement combinations. Separating out the values of the component events would be a more difficult, but not necessarily impossible, task.

The three combinations studied have been (*a*) RN—*E* says "Right" (R) following a correct response and nothing (N) following an incorrect response, (*b*) NW—*E* says nothing following a correct response and "Wrong" (W) following an incorrect response, and (*c*) RW—*E* says "Right" following a correct response and "Wrong" following an incorrect response. In these experiments the RN combination has been shown to be less effective in producing learning than the other two combinations, while the latter two have been about equally effective.

To explain these results, Buss, Braden, Orgel, and Buss (1956), and Buss, Wiener, and Buss (1954) have postulated an underlying reinforcement continuum on which N occupies a point of no reinforcing effect with R lying a short distance away on the positive side and W a much greater distance away on the negative side. In other words, W is assumed to have a greater effect on response probability than R. Earlier, a formulation in which N was assumed to have a weak positive reinforcing effect was considered but this was later dropped because it was felt not to be consistent with, ". . . known facts concerning extinction" (Buss et al., 1956, p. 289).

Consideration of the experimental situation and the obtained data suggests an alternative explanation of these results. Assume that during

[1] The author wishes to gratefully acknowledge the assistance of Solomon Feldman and Ann Sowards in collecting the data.

the course of experimental trials N acquires a negative reinforcement value in the RN combination and a positive reinforcement value in the NW condition. These values could be acquired through contrast of N with R, or with W. If it is further assumed that N acquires a greater (absolute) reinforcement value in the NW combination than in the RN combination, or acquires its value more rapidly in the former condition, the data of Buss et al. can be explained without recourse to the assumption of unequal values for R and W. The postulated difference in the reinforcement value of N in the two combinations could result from N having an initial positive value, from differences in the proportions of trials on which N occurs under the two conditions, or from differences in *S*s' reactions to the absence of R as opposed to the absence of W.

The present paper is concerned with a test of the assumption that N acquires reinforcement value during the course of acquisition. Consider an *S* who has undergone a series of acquisition trials under either RN, NW, or RW conditions which is followed by a series of trials on which N follows all responses (extinction). If acquisition has taken place under RN, N will be a negative reinforcer but not otherwise. Hence, it can be predicted that extinction will occur more readily following acquisition under RN than under the other conditions.

The experiment described below was intended to test this hypothesis. Additionally, it was desired to test whether the results of Buss et al. could be duplicated using a different learning task and specifically one which was suitable for further experiments which had been planned. The learning task utilized is partially designed for the requirements of these additional experiments.

The task, which might be called "pseudo-coding," requires *S* to select one member of a pair of nonsense syllables which are presented jointly with a three-digit number. Despite the somewhat misleading instructions given to *S*, the numbers were merely irrelevant stimuli. In essence, *S* was presented repeatedly with each of three pairs of nonsense syllables and required to select one, with reinforcement contingent only on which syllable was selected, i.e., one member of each pair was always correct for a given *S* regardless of the number on the card.

Method

Subjects.—The *S*s were 54 undergraduates enrolled in introductory psychology sections.

Materials.—Packs of 3 × 5-in. white cards were prepared, each with a three-digit number typed on the left side and a pair of nonsense syllables typed one above the other on the right side. Three syllable pairs were used, each syllable appearing in the upper position on half the cards on which that pair appeared. Within a pack each number appeared with only one syllable pair, three different numbers appearing equally often with any one pair, and with the two syllable positions for that pair. Thus there were 18 nonidentical cards in each pack. A pack consisted of 216 cards arranged in a random order with the restrictions that each block of three cards contain the three different syllable pairs and that each block of 18 cards contain no two identical cards. There were three such packs, differing only in which numbers occurred with a given syllable pair, and in card orders.

All syllables used were obtained from the list of 40% association-value syllables prepared by Glaze (see Hilgard, 1951). They were selected randomly but with restrictions which served to minimize reoccurrence of vowels and eliminate reoccurrence of consonants from the six syllables. One syllable in each pair was randomly assigned to Response Class 1, the other to Response Class 2. Numbers used were selected from a table of random numbers and randomly assigned to syllable pairs for the first card pack. For additional packs, numbers and syllables were assigned to each other to form a Latin square.

Procedure.—Each *S* was run individually. The *S* sat facing *E* across a table and was read the following instructions:

"This is a coding task. You will be shown a series of cards one at a time. Each card has a three-digit number and two three-letter syllables on it. One of the syllables on each card goes with the number on that card. You don't know, now, which syllable goes with a particular number, but it will become clear to you as we go. Each time I show you a card I want you to spell out the syllable which you think goes with the number. To begin with you may have to guess but as we go along it will become clear to you which syllable goes with the number on the card."

Questions about how *S* would know which syllable went with a number were answered by *E* saying, "It will become clear to you as we go."

Following the instructions the cards were exposed one at a time. When *S* responded, *E* either said "Right" or "Wrong," or said nothing (whichever was appropriate), recorded the response and exposed the next card.

Eighteen *S*s were given each of the following treatments for 144 acquisition trials: (RW) *E* said "Right" for an A-response, "Wrong" for a B-response; (RN) *E* said "Right" for an A-response, nothing for a B-response; (NW) *E* said nothing for an A-response, "Wrong" for a B-response. Following acquisition, with no break in the procedure, *E* said nothing following every response for 72 trials. This "extinction series" was identical for all *S*s.

The *S*s were assigned to card packs, response classes, one of two *E*s, and experimental conditions, by a random assignment method subject to restrictions which balanced packs, response classes, and *E*, in the experimental groups and resulted in each *E* running the same number of *S*s under each combination of packs and responses in one group as he did in any other group. Assignment to a response class meant that responses in that class were followed by the reinforcement event assigned to A, responses in the other class being treated as instances of Response B (see above).

Results

Acquisition.—Tabulations of total number of B-responses during acquisition, and of the last acquisition trial on which a B-response occurred, were made for the three experimental groups. Summary data are shown in Table 1. Since the distributions were badly skewed, over-all analyses were made using the Kruskal-Wallis (1952) test. For last trial on which a B-response occurred, the three groups differ significantly at the .05 level ($\chi^2 = 7.30$ with 2 *df*). For number of B-responses the results barely fail to attain the .05 level ($\chi^2 = 5.77$, as compared with 5.99 for $P = .05$ with 2 *df*). Distributions of scores for all pairs of groups were tested using the Mann-Whitney test. For both last trial and number of B-responses, only the RW-RN difference is significant at the .05 level.

TABLE 1

Summary Statistics of Acquisition Data (Replication I)[a]

Group	Number of B-Responses			Last Trial on Which B-Response Occurred		
	Mean	Mdn.	*SD*	Mean	Mdn.	*SD*
RN	20.56	15.0	17.87	63.28	45.5	49.66
NW	13.61	11.0	9.88	43.78	28.5	40.16
RW	11.78	5.5	16.37	28.56	12.0	37.05

[a] Data based on 144 trials per *S*, 18 *S*s per group.

Since it is impossible to conclude from these data that the effects of the NW condition differ from those of either the RN or RW conditions and since it cannot be equal to both, a replication of the entire experiment (using a new *E*) was accomplished with the aim of subjecting data from both replications to an analysis of variance (using a square-root transformation to normalize the data) in order to obtain a more powerful test of differences. The data from this replication are summarized in Table 2. Although the relative order of means is the same for this as for the earlier replication, neither measure shows significant variation among groups by the Kruskal-Wallis test.

TABLE 2

SUMMARY STATISTICS OF ACQUISITION DATA (REPLICATION II)[a]

Group	Number of B-Responses			Last Trial on Which B-Response Occurred		
	Mean	Mdn.	*SD*	Mean	Mdn.	*SD*
RN	26.72	27.0	18.83	67.28	62.0	44.05
NW	23.00	14.0	21.09	57.94	44.0	43.53
RW	18.17	10.5	17.38	44.22	33.0	42.82

[a] Data based on 144 trials per *S*, 18 *Ss* per group.

Two-way analyses of variance (Groups and Replications) of transformed scores yielded only two significant *F* ratios: between Replications in the number of B-responses data, and between Reinforcement Combinations in the last trial data. The within-groups variance estimate from the latter test yields *t* values which are significant at or beyond .05 in all three comparisons. Thus RW conditions result in quicker learning than NW conditions, which result in quicker learning than RN conditions, by this measure.

Extinction.—Examination of the extinction data from the first replication showed that 15 *Ss* in the RN group, 2 *Ss* in the NW group, and 1 *S* in the RN group, made at least one B-response during the 72 trials under "extinction" conditions. Results of the second replication were essentially similar, the number of *Ss* in each group who showed some evidence of extinction being 13, 1, and 3, for Groups RN, NW, and RW, respectively. Applications of the χ^2 test to these frequencies showed the differences to be significant beyond .001 for each replication.

At the end of the series of acquisition trials some *Ss* were still making B-responses. To test the hypothesis that extinction differences resulted from differences in the extent to which *Ss* in the various groups had acquired the A-response, the following two procedures were carried out. First, all *Ss* who made no B-responses on the last 36 acquisition trials were considered. A count was made of the number of such *Ss* in each group who showed some evidence of extinction, and of the number who did not. These data are shown in Table 3. The differences were evaluated by the χ^2 test for each replication separately. Both yielded highly significant χ^2 values ($P < .001$). Second, all *Ss* whose last B-response did not occur on a trial later than the median such trial in the RN group (using a separate median for each replication)

TABLE 3

NUMBER OF *Ss* WHO MADE B-RESPONSES DURING EXTINCTION FOR SELECTED GROUPS OF *Ss*

Group	*Ss* with 100% A-Responses On Trials 109–144				*Ss* with Last B-Response No Later Than Median of Group RN			
	Replication I		Replication II		Replication I		Replication II	
	Some B	No B	Some B	No B	Some B	No B	Some B	No B
RN	11	3	9	5	6	3	6	3
NW	1	15	0	15	0	11	0	12
RW	1	16	1	15	1	14	1	13

were considered. Extinction data for these *S*s were treated in a manner similar to those described for the first procedure. The data are shown in Table 3. Using the hypergeometric distribution (Feller, 1950), the probability of obtaining the frequencies found in the first replication is .0003 on the hypothesis that the proportion of *S*s who show some extinction is equal for the three groups. Since there are only a few other combinations which are less probable, and which satisfy the restrictions of the marginal frequencies, it is apparent that each of these sets of values has a $P < .01$ of occurring, under the null hypothesis.

Discussion

The present finding that, during acquisition, *S*s learning under RN tend to continue making B-responses (errors) longer than *S*s learning under NW is in agreement with the results of earlier experiments (Buss & Buss, 1956; Buss et al., 1954; Buss et al., 1956). The finding of differences in this respect between the NW and RW groups is different from previous findings although it is to be noted that the data from one experiment in which data are reported in terms of trials to criterion, and of number of *S*s still making errors at the end of 100 trials, seem to agree fairly well with the present acquisition results (Buss & Buss, 1956, Exp. 1). Since the present experiment differed from earlier ones in regard to type of task used, number of acquisition trials, and the population from which *S*s were drawn (the studies of Buss et al. largely used psychiatric hospital patients as *S*s), and since different data-analysis techniques were used, it is impossible to decide what is responsible for the difference in findings concerning acquisition under NW and RW.

It should be noted that the acquisition results of this experiment are predictable from either the assumption of unequal reinforcement values for R and W, or from the assumptions made earlier in this paper. So long as R is assumed to be more positively reinforcing than N, the RW combination must provide more over-all reinforcement than the NW combination.

The extinction data are not predictable from both sets of assumptions. If N has no reinforcing effect, or the same reinforcing effect regardless of acquisition conditions, there is no basis for predicting differences in the extinction series. On the other hand, the assumption that N becomes a negative reinforcer during acquisition under RN conditions, a positive reinforcer under NW conditions, and has its initial value unaltered under RW conditions enables the extinction findings to be predicted.

Summary

College students were given 144 trials of learning on a task requiring *S* to spell one of a pair of nonsense syllables presented to him. The *E* said "Right" following an A-response and nothing following a B-response; nothing following an A-response and "Wrong" following a B-response; "Right" following an A-response and "Wrong" following a B-response for *S*s in the RN, NW, and RW groups, respectively. On 72 additional trials *E* said nothing regardless of *S*'s response.

Acquisition of the A-response occurred most rapidly in the RW group followed by NW and RN in the order. When N followed all responses the amount of extinction of the A-response was found to be greatest for the RN group. For the majority of NW and RW *S*s, B-responses do not occur on extinction trials. Further data analyses indicate that extinction differences are not attributable to differences in extent of acquisition. The results were interpreted as confirming the assumption that N becomes a negative reinforcer during acquisition under the RN condition.

References

Buss, A. H., & Buss, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283–287.

Buss, A. H., Braden, W., Orgel, A., & Buss, E. H. Acquisition and extinction with differ-

ent verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

Buss, A. H., Wiener, M., & Buss, E. H. Stimulus generalization as a function of verbal reinforcement combinations. *J. exp. Psychol.*, 1954, **48**, 433–436.

Feller, W. *An introduction to probability theory and its applications*. New York: Wiley, 1950.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Kruskal, W. H., & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. statis. Ass.*, 1952, **47**, 583–621.

(Received February 20, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

PERFORMANCE DECREMENT AT AN AUDIO-VISUAL CHECKING TASK [1,2]

WILLIAM E. KAPPAUF

University of Illinois

AND

WILLIAM E. POWE

Air Force Personnel and Training Research Center

This is a study of performance change during long periods of observation. Experiments aimed at securing a better understanding of such changes are not uncommon in our recent literature and reflect growing interest in the problem of maintaining the alertness or vigilance of observers, drivers, and other workers at their jobs.

The particular task which *Ss* in the present experiment were required to perform was an audio-visual checking task. Each *S* listened to a long, continuous series of digits presented by tape recorder and checked these numbers against an almost identical series presented in a mimeographed test booklet. He worked for 2 hr. without interruption, marking his test booklet whenever he detected a discrepancy between the two series.

The purposes of the study were to determine whether performance at this task would change with time, and whether such changes as might be observed would be related to (*a*) rate of occurrence of discrepancies in the digit series, or to (*b*) aptitude level of *S*.

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726 monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part, by or for the United States Government. This paper is a revision and consolidation of two reports transmitted by the Psychology Department, University of Illinois, to the sponsoring agency under the following titles: "Performance decrement at an audio-visual checking task," 28 February 1955; "Relation of certain aptitude scores to performance decrement at an audio-visual checking task," 14 October 1955.

[2] The authors acknowledge the assistance of John D. Davis and Henry F. Gromoll in the analysis of the data of this experiment.

PROCEDURE

Test materials.—The tape recording provided five series of digits: four short practice series, and the 2-hr. test series. All digits were presented at the rate of 1 per sec. The sequence was monotonous and without inflection, having been prepared by splicing together reproductions of single original recordings of the digits from one to nine.

The visual digit series was contained in a mimeographed test booklet consisting of two practice pages and 15 test pages. The digits on each page were double-spaced within and between rows. Nine and two-thirds pages were required to cover the 2-hr. test, but additional pages were included in each booklet so that *Ss* would not know precisely when the task would end. There were four sets of test booklets, all with standard practice pages but with different test pages. In one set, the latter had been prepared with 2 randomly located discrepancies in each 15-min. section. In the three other sets of booklets, the test pages contained 5, 10, and 20 discrepancies, respectively, in each 15-min. section. No given discrepancy was common to two or more series.

Thus, discrepancy rate was determined by the test booklet with which *S* was provided, and different *Ss* could be tested simultaneously under different discrepancy conditions.

Aptitude measures.—Each *S*'s category score on the Armed Forces Qualification Test (AFQT) was available prior to the experiment. This test provides a composite ability score representing performance on items which involve arithmetic reasoning, word knowledge, knowledge of

tool functions, and spatial visualization. Scores are recorded by category, and placement in one of the first four categories is required for service in the Air Force. Norms for the entire military population, officers and men, define these categories in terms of percentiles in the following way: Category 1, 93 to 100; Category 2, 65 to 92; Category 3, 31 to 64; Category 4, 10 to 30.

Design.—The design was factorial: four categories or levels on AFQT, by four discrepancy conditions.

Subjects.—The *Ss* were 235 basic trainees at Lackland Air Force Base. All were in their second week of training. They were from four training flights of approximately 60 each. Thirty men were tested at a time—half of a given flight in the morning, the remainder in the afternoon. Every group contained men from all four AFQT categories. Assignment of men to discrepancy levels was controlled so that, on a running basis throughout the experiment, approximately equal numbers from any one AFQT category were tested under each of the four discrepancy conditions. That the final balance was good may be seen from the notations on group size given in Fig. 1 and 2.

Testing conditions.—The rooms used for testing were divided into small cubicles. Each cubicle was the work space for one *S*. The *Ss* in adjacent cubicles worked under different discrepancy conditions and because discrepancies in the different booklets never coincided, there was no opportunity for *S* to be "tipped off" to a discrepancy by the detection responses of his neighbors. All *Ss* worked facing the wall. A loudspeaker from the tape recorder was located at the door of each test room.

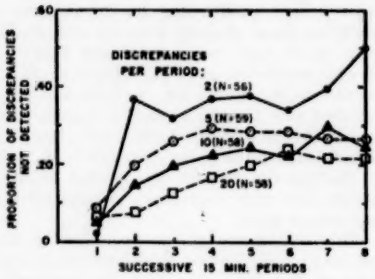

FIG. 1. Proportion of discrepancies not detected as a function of working time, for each of the four discrepancy rates. Data for men in different AFQT categories have been pooled in obtaining each of the four curves. The *N*'s indicate the distribution of the 231 *Ss* across the four tasks.

Instructions and practice.—The *S*'s were given this test as if it were part of the Air Force airman classification test program and were advised that it would be in their best interest to do as well as they could at all times. Their specific work instructions appeared on the cover page of the test booklets and were read aloud to them. These began:

"Your task in this experiment is a proofreading task. You are to check whether the numbers on your pages agree with the numbers that you hear over the loudspeaker. Whenever you find a number which does not agree, you have detected a typing error. Cross out the wrong number. For instance, if you hear 1 4 6 6 9 and your answer sheet reads 1 4 7 6 9, draw a line through the 7."

The instructions further indicated that the work would go on continuously from each page to the next, and that each *S* was to mark his booklet immediately with a check if he should lose his place (as evidenced by two or more successive apparent discrepancies) and mark it with an "S" as soon as he recovered and was starting to work again. The need for close attention to the task was emphasized.

The practice series were then presented. The first three were of 2-min. durations and contained 10 discrepancies each, while the last was of 5-min. duration with 8 discrepancies. After every series, each *S* scored his own performance and questions on procedure were answered.

Work on the test series began following a 15-min. "break" and a brief additional instruction which noted that the number of wrong numbers found during the practice trials was completely unrelated to the number that might be found in the test itself. The *Ss* were not told how long the test would last, but they probably anticipated that it would conform to their usual morning or afternoon duty schedule. They were not required to give up their watches during the test.

Mastery of the task.—Detection of all discrepancies in at least one of the four practice periods or in the first 15-min. period of the test was taken as evidence that *Ss* understood the task they were to perform. Of 235 men who were tested, 231 met this criterion. The data reported below are for the later number. Of these 231, 165 detected all discrepancies in the first 15 min. of the test.

RESULTS

Principal attention is given here to the occurrence of omission errors, i.e., failure of *S* to detect discrepancies.

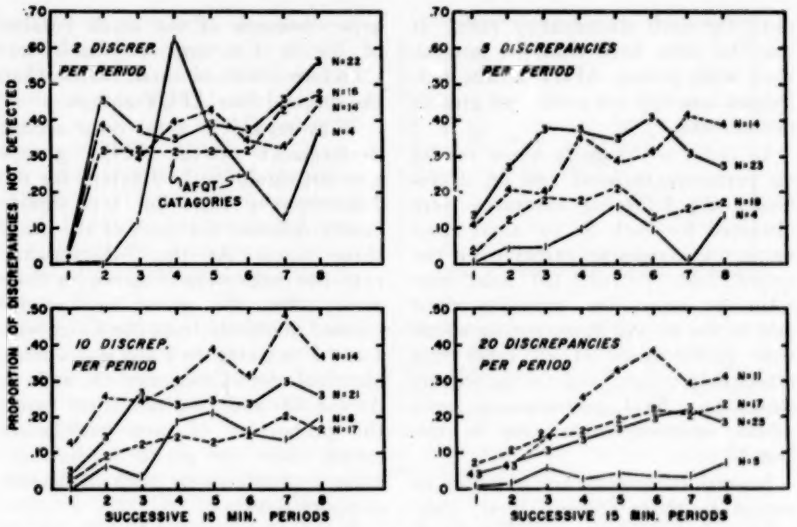

FIG. 2. Comparison of performance of men in the four AFQT categories at each of the discrepancy rates, in terms of the proportion of discrepancies not detected as a function of working time. The small numbers which code the curves identify the AFQT categories. The *N*'s indicate the distribution of *S*s across categories.

Brief comment on commission errors, i.e., discrepancy responses made to digits which were really correct, and on reports of lost places appears at the end of this section.

Omission Errors

Time and discrepancy rate.—The record of all omissions for all *S*s is summarized in Fig. 1. Each curve in this plot is for one discrepancy condition and is based upon data pooled across AFQT levels. Note that the ordinate represents the proportion of discrepancies which were not detected. The data in the figure may be summarized as follows: (*a*) There was a performance decrement at each of the four tasks. (*b*) Performance level following decrement was progressively poorer for the lower discrepancy rates. (*c*) The greater the decrement, the more rapid the rate at which it developed.

Because successful performance at this audio-visual checking task depended on *S*'s keeping his place in the number series, a check was made to determine whether the results reported here were seriously influenced by *S*'s getting lost. A second set of omission error scores was obtained, taking into account only those discrepancies which occurred when there was no notation on the booklet to imply that *S* had lost his place. This change in computation, which operated to reduce all the omission percentages, lowered the curves in Fig. 1 slightly but did not change the shapes of the curves or the relations between them. Thus, insofar as the lost-place reports were accurate, the results shown in Fig. 1 apply equally well to detection when *S* was not lost in the number sequence.

AFQT score.—Figure 2 presents a breakdown by AFQT category of the

data for each discrepancy rate. It may be seen here that, in general, men with poorer AFQT scores performed less well and evidenced greater decrements.

In order to quantify these results on performance level and on decrement, the following measures were obtained for each *S*: (*a*) total error score, i.e., omission errors over the entire 2-hr. period; (*b*) final performance score, i.e., omission error rate in the second hour, during which time performance at all tasks was reasonably stable; and (*c*) decrement score, i.e., final performance score minus omission error rate in the first 15 min.

Several comments are in order in regard to these scores. First, they are not independent, but the use of all three of them appears justified because of the different ways in which other authors have reported similar data. Secondly, their reliability is expected a priori to vary with the discrepancy rate under which *S* was tested, because the number of observations contributing to each score increases with discrepancy rate. Thirdly, the reliability and the distribution of the decrement scores are necessarily influenced by the low reliability and the exceptionally skewed distribution of the error scores for the first 15 min., which conditions suggest that it would be inappropriate to try to develop a more elaborate assessment of decrement as, for example, in terms of final performance scores adjusted by covariance methods for initial performance.

The method chosen for evaluating the foregoing measures was to use .05 level median tests for the effect of AFQT category. For this purpose, *S*s with AFQT category scores of 1 and 2 were combined into a single group because of the small number of *S*s in Category 1. Each test therefore involved three rather than the original four AFQT groups.

With regard to total error scores, performance of the AFQT groups was not significantly different for the 2-discrepancy rate, but was significantly different for each of the other three rates. At the 5-discrepancy rate, the proportion of men with more errors than the grand median increased markedly from the Categories 1 and 2 to Category 3 and was almost identical for Categories 3 and 4. At the 10- and 20-discrepancy rates, the proportion of men with more errors than the grand median increased progressively with increasing category scores.

With regard to final performance scores, the proportion of men with more errors than the grand median increased progressively with increasing category number in the contingency table for each discrepancy rate. Differences between AFQT groups were significant for the 5- and 10-discrepancy conditions and approached significance for the 20-discrepancy condition where $.10 > P > .05$.

With regard to decrement scores, the proportion of men with larger scores than the grand median again increased progressively with increasing category number in the contingency table for each discrepancy rate. Performance of the AFQT groups was significantly different under the 10- and the 20-discrepancy conditions. The magnitude of the group differences in the two latter conditions is best shown by a comparison of the average decrement for the men in Categories 1 and 2, which was 13.3%, and the average decrement for the men in Category 4, which was 26.7%.

Alternatively, the reader might be

interested in correlation measures of the relation of decrement to AFQT. The Pearson r's, which will here be attenuated because of the sizeable grouping errors on AFQT, were as follows: +.30 for AFQT and decrement score for the 20-discrepancy condition, +.21 for the 10-discrepancy rate, +.18 for the 5-discrepancy rate, and +.07 for the 2-discrepancy rate. Of these, only the first is significantly different from zero. For purposes of later discussion, it should also be added that a test of the hypothesis that these four correlations are homogeneous leads to the acceptance of that hypothesis.

These analyses of the data represented in Fig. 2 may be summarized by saying that the relation of performance and of decrement to AFQT seen in the figure proves to be statistically significant for the higher discrepancy rates where the checking test scores are the most reliable.

Time since previous discrepancy.—There has been general interest in the problem of determining whether the probability of S's detecting a signal varies with the amount of time which has elapsed since his last detection. Mackworth (1950), Deese and Ormond (1953), and Jenkins (1953) have all inquired into this matter with results which could hardly be more contradictory. Mackworth's data imply improved probability of detection as the interval increases. Deese and Ormond report no statistically significant dependence of probability of detection upon the time interval. Jenkins notes a different relationship for two different rates of signal presentation, and the relation which he presents for the high signal rate of 120 signals in 15 min. is just the opposite from Mackworth's trend. These differences in results may mean either that the probability of detection is basically unrelated to the intertrial interval, or that the relation is critically influenced by conditions and schedules specific to the individual experiments. In either event the matter deserves further study and so the present data were analyzed in terms of the time variable.

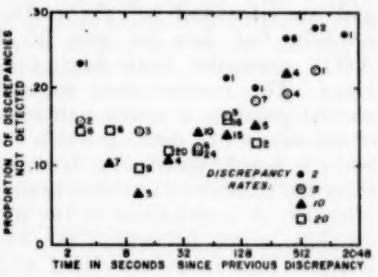

FIG. 3. Proportion of discrepancies not detected as a function of time since previous detected discrepancy. Data for men in all AFQT categories have been pooled. Plotted points are coded for the discrepancy rates which they represent. The number next to each point identifies the number of discrepancies (or trials) for which data were available in computing the indicated proportion.

At the start of the analysis, two steps were taken to reduce unwanted variance: (*a*) The data which were used were limited to plateau regions on each of the decrement curves. Mackworth limited his data in a similar way. (*b*) The trials which were admitted to the analysis were limited to those where S had in fact detected the previous discrepancy and had made no commission errors since that last detection. Deese and Ormond employed a similar control, as did Jenkins.

The obtained data are summarized in Fig. 3. Here the probability of nondetection is plotted as a function of time (log scale) since last detection. Data for trials which occurred under each of the four discrepancy conditions have been plotted separately

and are identified by the plotted symbols, but data for men in all AFQT categories have been combined. The number next to each plotted point is a rough indication of the amount of data on which the point is based: specifically, it represents the number of interdiscrepancy intervals, as programmed in the test booklets, which fell within the logarithmic time range associated with that point. Note that these numbers vary from 1 to 24.

Unfortunately there is no fully appropriate statistical test to apply to data of this sort where for a single point some *S*s are represented several times and others less often or not at all (due to data elimination described above), and where it is further, known that probability of detection varies for different *S*s. For want of a proper test, tests using binomial probability paper have been applied in the hope that they will not be too misleading. Under this test, the plotted probabilities for the 10-discrepancy condition vary significantly (.05 level test). So also do the points for the 20-discrepancy condition.

The data in Fig. 3 suggest that the relation of the probability of detection to time between trials may be U-shaped. The U-shaped trend appears in the results for the 5- and the 10-discrepancy rates individually. How much the over-all pattern as seen in the figure is exaggerated by the points for the 2-discrepancy condition, for which no observations were available between 4 and 64 sec., could only be determined from additional and more complete data. One weak test on the present data is possible, however, and that is a test of whether the points at a given time-between-discrepancies are homogeneous over discrepancy rates. These differences have been tested on binomial probability paper. Only one set of values, those for the 8- to 16-sec. range of times, involves differences which even approach the .05 level of significance according to this test.

The implication of the foregoing analysis is that detection probabilities were comparable at different discrepancy rates but did vary with the length of the interdiscrepancy interval.

Commission Errors and Lost Places

Commission errors in this and other vigilance situations are frequently a function of *S*'s standard or criterion of what constitutes an event of the sort for which he is looking. If the situation entails ambiguity, some *S*s "play it safe" and in so doing make many commission errors. It was not intended that ambiguity be present in the checking task used in this experiment. Some arose inadvertently, however, due to variations in the quality of the mimeographing in small regions on some pages. This led to an unfortunate contamination of the commission error data and so they have not been analyzed.

The number of reported lost places was not great for any of the experimental conditions. When the record of these reports was analyzed, their frequency was found to be unrelated to discrepancy rate or to AFQT. The trend over the 2-hr. work period was for the occurrence of reports to increase to the third 15-min. period and decline thereafter.

Discussion

The general pattern of results in Fig. 1 is similar to that reported by Jenkins (1953) for a study in which *S*s monitored periodic deflections on a voltmeter for 90 min. and were required to report

the occurrence of supernormal deflections. Jenkins, like the present writers, followed Mackworth's (1950) clock test procedure in using a regular, unchanging stimulus rate of one event per second. Most swings of the needle were of standard amplitude. Supernormal swings occurred at rates of 480, 60, 30, and 7.5 per hour for four different groups of *S*s, all enlisted Air Force personnel. Omission errors were more frequent in Jenkins' experiment as compared with the present one, but in both experiments, performance at the several signal rates was described by a family of decrement curves.

Deese and Ormond (1953) present performance curves for radar search with 40 target presentations per hour and 10 presentations per hour. Poorer performance was associated with the lower rate, in agreement with Jenkins' data and those of the present study. Deese and Ormond did not, however, obtain curves showing decrements resembling those of Fig. 1. Possible reasons for this could be several, among them their summarizing of the data by half-hour periods which might have obscured more proficient early performance.

The results cited above on the relation between *S*s' general aptitude test scores and their performance decrements are in accord with results recently reported by Broadbent (1954) for a small number of *S*s, as well as with data which were collected during exploratory work with the present audio-visual checking task. Subjects used in this pilot work included 36 college students and 16 airmen stationed at Chanute Air Force Base. The discrepancy rate was 6 in 15 min. Over the first hour of testing, the decrement in performance of the more highly selected college group was less than half that of the airman group.

The obtained relation of decrement to aptitude is interesting because it implies a selection and placement strategy directly opposite to the oft-quoted one of assigning dull men to dull jobs. Whether the effective selection of men for jobs requiring vigilance ever develops from this lead depends upon answers to at least two critical questions. The first is whether the present relationship has resulted from a factor such as test wiseness or ego involvement common to performance at both the AFQT and the checking task. The checking task was, after all, presented as a test in a test situation and with presumed consequences for each man's military future. Only if this test character of the checking task proves unimportant for the obtained correlation may the relation of performance decrement to aptitude be expected to apply in actual duty situations. The second question is whether the decrement differences between men of different aptitude levels will become more systematic at low discrepancy rates if the decrement measures at these rates are made more reliable, as by averaging the performance of individual *S*s over many test sessions. The correlations of decrement score and aptitude for the four discrepancy conditions were noted above to be homogeneous, but the suspicion is strong that the trend of the correlation values is real and is associated primarily with the reliabilities of the several sets of scores. Obtaining adequate data for the case of low signal rates will not be an easy matter, but it is the important case because of the fact that most watchkeeping jobs, by their very nature, entail very low rates of signal presentation.

The present correlations do contain a practical suggestion for future studies of vigilance, however, and that is that it would be wise to avoid the exclusive use of *S*s of high general aptitude.

Interpretation of the results for interdiscrepancy interval is difficult at this time. There are many differences between the several experiments mentioned earlier as having explored this interval problem. The nature of the task and of the *S* populations are two dimensions of difference. The pattern of the times between trials is another. Mackworth's times between trials were patterned in the same way every 30 min., those in the present experiment were not. Deese and Ormond thought it best to eliminate variation in the frequency of occurrence

of different time intervals. They used a rectangular distribution of times, with a maximum time of 5 min. But this distribution, too, may have influenced their *S*s' behavior, in that *S*s may have learned reasonably early in the tests that delays of more than 5 min. never occurred. The range of intertrial intervals must also be of some importance in determining the picture which is found. To date, the present study reports the widest range of intervals.

Clearly needed are studies which will look directly and primarily at this interval question. Perhaps the time distribution of trials which should receive principal consideration is that where the probability of a signal occurring in any short time interval, say 2 sec., is always the same. This would produce a J-shaped distribution but it would be one in which *S* could never build up any valid expectancies based on how much time had elapsed since he last detected a signal.

Summary

Two hundred and thirty-five Basic Airmen, distributed over all four scoring categories on the Armed Forces Qualification Test (AFQT), were tested on an audio-visual digit checking task which continued for 2 hr. Discrepancies which *S*s were to detect between the audio- and the visual-digit series occurred at four different rates for different groups. These were 2, 5, 10, and 20 discrepancies per 15-min. period of the test. The data which are reported are for 231 *S*s who met an initial mastery criterion.

Performance decrements over the 2-hr. period were found for all AFQT category groups and for all four discrepancy conditions. Differences between conditions and between AFQT groups were found. Greater decrements occurred for the slower discrepancy rates and, in general, for men of poorer AFQT score.

The relation between the probability of detection of a discrepancy and the time since the previous discrepancy was found to be U-shaped.

References

Broadbent, D. E. Some effects of noise on visual performance. *Quart. J. exp. Psychol.*, 1954, **6**, 1–5.

Deese, J., & Ormond, E. Studies of detectability during continuous visual search. *USAF WADC Tech. Rep. 53–8*, September, 1953.

Jenkins, H. M. Performance at a visual monitoring task as a function of the rate at which signals occur. *Massachusetts Institute of Technology Tech. Rep. 47*, 29 October 1953.

Mackworth, N. H. Researches on the measurement of human performance. *Med. Res. Council, Spec. Rep. Ser. No. 268*. London: His Majesty's Stationery Office, 1950.

(Received March 12, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

PERFORMANCE IN DIFFERENT SEGMENTS OF AN INSTRUMENTAL RESPONSE CHAIN AS A FUNCTION OF REINFORCEMENT SCHEDULE [1]

KENNETH P. GOODRICH

State University of Iowa

In a review of the literature up to 1950, Jenkins and Stanley (1950) included instrumental reward conditioning in their general conclusion that continuous reinforcement leads to superior acquisition performance as compared with partial reinforcement. Although there is some evidence (Finger, 1942a, 1942b; Lewis, 1956; Lewis & Cotton, 1957; Sheffield, 1949; Wilson, Weiss, & Amsel, 1955) in support of this conclusion, the generality of these findings is made questionable by the fact that these studies employed, at most, only 30 acquisition trials.[2] Quite a different picture is suggested by the studies of Weinstock (1954) and Haggard (1956) which employed a considerably more extended acquisition period. Their findings indicate that (*a*) the initial decrement in performance under partial reinforcement is temporary and (*b*) in the case of some response measures at least, the final or asymptotic level of performance may be greater under partial than under continuous reinforcement.

While Weinstock did not find a significant advantage in the early period of training for the continuous group, or a consistent and significant advantage for his partial groups in the later stages of training, the acquisition curves of running speed did indicate an interesting pattern of initial divergence, convergence, and finally a reversal of the initial order among three of his four groups. Following up this suggestive finding, Haggard obtained further evidence which showed the same pattern (initial advantage for the 100% group and subsequent reversal) for curves of starting speed. In the case of these data the final superiority of the 50% group over the 100% group was significant at the .05 level. In the case of a running speed measure, Haggard found an initial advantage for the 100% group (significant at the .10 level) with the two curves finally converging at a very late stage of training. This finding suggested the possibility that the effects of variation of the reinforcement schedule in such an instrumental conditioning situation may depend upon which portion of the response chain is measured. The recent finding of Freides (1957) which showed that speed of running in the goal box end of the runway was faster for 100%-reinforced *Ss* than for 50%-

[1] Part of a thesis submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the M.A. degree. The author is indebted to Kenneth W. Spence for advice and assistance throughout the course of the investigation. The present report was prepared while the author was a National Science Foundation Predoctoral Fellow.

[2] Four of these studies (Lewis, 1956; Lewis & Cotton, 1957; Sheffield, 1949; Wilson, Weiss, & Amsel, 1955) introduced a complication by administering 10 continuously reinforced pretraining trials in the experimental apparatus prior to the 30 subsequent acquisition trials. It is difficult to say what subsequent effects upon the relative performance of the two groups these trials may have had.

reinforced *Ss* lends further support to this suggestion.

The present investigation was undertaken to obtain further information on the effects of partial reinforcement in the runway situation, particularly with respect to the questions of (*a*) whether the effects will vary at different points of the response chain and (*b*) whether, for some response measures at least, 50%-reinforced groups will respond faster at the asymptote of training than 100%-reinforced groups. The latter question has important implications for the frustration theory of Amsel (1958) and for the statistical-contiguity learning theory discussed by Weinstock (1954).

Experiment I

Method

Subjects and apparatus.—The *Ss* were 30 experimentally naive female hooded rats from the colony maintained by the Psychology Department of the State University of Iowa. Ages ranged from 114 to 149 days at the beginning of the experiment.

The apparatus was a straight alley consisting of an 8-in. start box, a 13-in. alley, and a 12-in. goal box. These sections were all 2 in. wide and were covered with one continuous sheet of clear glass at a height of 3 in. Two start doors, one opaque and the other clear plastic, separated the start box from the alley. An opaque retrace door separated the alley from the goal box. These doors slid horizontally through slots in the walls. The interior of the apparatus was painted flat black, and the floor was covered with black rubber matting. Dim illumination was provided by the reflected light from two shielded 200-w. bulbs located 4 ft. overhead.

Starting, running, and goal box time measures were provided by three timing circuits activated by the interruption of infrared photocell beams located 6 in., 12 in., and 24 in., respectively, beyond the starting doors. The last beam was 1 in. from the end of the goal box and went across the top of the food cup.

Drive maintenance and pretraining.—During the 10 days prior to the first acquisition trial the *Ss* were habituated to handling and a 23-hr. hunger drive cycle was established. Each *S* was fed 1 hr. daily in an individual feeding cage. Fresh water was always available in the home and feeding cages.

Pretraining began 6 days prior to the first acquisition trial. On Day 1, *Ss* ate for 30 min. in the feeding cages from a supply of the experimental 45-mg. pellets (Noyes Co.). On Days 2–6, *Ss* received pretraining feedings in a separate gray goal box in the experimental room. Day 2 consisted of one such 8-pellet feeding, and Days 3–6 consisted of two 4-pellet feedings each day.

Experimental training.—Following pretraining the *Ss* were randomly assigned to two groups of 15 animals each. The *Ss* in the 100% group were reinforced on all trials with four of the 45-mg. pellets. The *Ss* in the 50% group were nonreinforced on two trials of every block of four trials, each block consisting of one of the following patterns: +−+−, −+−+, +−−+, −++−. One trial per day was run on Days 1 and 2, two trials were run on Day 3, and four trials per day were run beginning on Day 4. Sixty trials were run in all.

On reinforced trials *Ss* remained in the goal box until they turned away from the empty food cup. On nonreinforced trials (no pellets in the food cup) each *S* remained in the goal box after breaking the last light beam for a period equal to that *S*'s eating time on the previous reinforced trial, or for 45 sec., whichever was shorter. The *Ss* were returned to individual waiting cages between trials. The intertrial interval was approximately 17 min.

Results

The three time measures were converted to speeds (reciprocal time). Two additional sets of observations were recorded: (*a*) frequency of occurrence of stopping responses made by *Ss* in each section of the apparatus and (*b*) the time spent by each *S* at the food cup, i.e., the time between *S*'s breaking the light beam over the loaded cup and his turning away from the empty cup.

Locomotor performance.—Starting, running, and goal box speeds are plotted in Fig. 1 for the two groups by blocks of four trials. In the starting and running speed segments of the alley (*a*) the 50% *Ss* ran slower than the 100% *Ss* in the early blocks of trials and (*b*) the 50% *Ss*

ran consistently faster than the 100% *Ss* during the later trials.

This reversal was expected, but prediction of the trial point at which it would occur had not been possible. Statistical evaluation of the effect must therefore depend to some extent upon the particular crossover point obtained in the data at hand. The most conservative procedure would seem to be a two-way analysis of variance with pre- and post-crossover means as the criterion measures. Such an analysis (Lindquist, 1953, Type I design) was computed for the starting and running speeds separately. If the crossover is reliable, a significant interaction of trials with reinforcement schedule should be found. This interaction was significant in both analyses ($F = 13.91$, $df = 1$ and 28, $P < .001$, and $F = 9.37$, $df = 1$ and 28, $P < .005$, respectively). Analyses were then computed on the pre- and post-crossover data separately. Prior to the crossover, the effect of reinforcement schedule was nonsignificant for both speed measures ($.10 < P < .20$ in both cases). After the crossover, the reinforcement effect was significant at between the .05 and .10 levels ($F = 4.06$, $df = 1$ and 28) for the starting speeds, but was nonsignificant for the running speeds ($P > .20$).

For the goal box speed measure, no crossover occurred, a fact consistent with Freides' finding with a similar measure. Because the two curves were so nearly parallel over the 60 trials, an analysis was computed using the mean speeds over all 60 trials. The difference between groups was significant at between the .05 and .10 levels ($F = 3.81$, $df = 1$ and 28). The goal box data for the 50% group were plotted for rewarded and nonrewarded trials separately. The observed differences were negligible and no statistical evaluation was carried out. It seems clear from this result that the relatively poorer performance of the 50% group in the goal box region did not result from these *Ss* slowing down on nonrewarded trials as they approached the food cup.

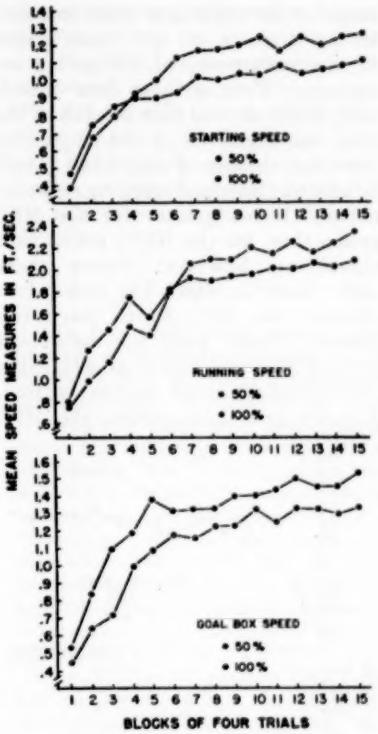

FIG. 1. Mean starting, running, and goal box speeds for the two groups in Exp. I by blocks of four trials.

An examination of the three sets of curves in Fig. 1 reveals a clear gradient effect. The 50% decrement is negligible in the first 6 in. of the alley, extends to 24 trials in the second 6 in., and extends over all 60 trials in the last 12 in.

Frequency of stops and time spent at the food cup.—Observations of

stops in the apparatus were recorded in an attempt to get some index of the occurrence of competing responses. Plots of these data (Goodrich, 1958) showed that in each of the first two segments of the apparatus the total number of stop trials (trials in which at least one stopping response occurred) was greater for the 50% group than for the 100% group (not significant, however) during those early trials on which the speed data showed the 50% *Ss* to have run slower. There were no differences beyond the trial points at which the corresponding speed curves crossed. For the goal box region, the 50% group made significantly more stops in the period up to Trial 20, at which point the curves converged and remained together.

On the two trials in each block on which the 50% *Ss* received food reward, the time spent at the food cup was recorded for both groups. Plots of the data (Goodrich, 1958) showed diverging negatively accelerated decreasing functions, with the 50% *Ss* staying consistently longer at the food cup than the 100% *Ss* throughout training. The groups differed significantly during trials 45–60 ($t = 2.27$, $df = 28$).

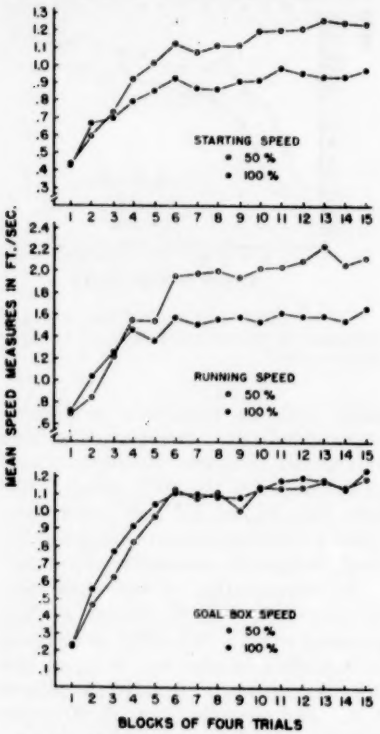

FIG 2. Mean starting, running, and goal box speeds for the two groups in Exp. II by blocks of four trials.

EXPERIMENT II

Method

Subjects and apparatus.—The *Ss* were 30 experimentally naive female hooded rats from the Iowa colony, ranging in age from 114 to 149 days at the beginning of the experiment. The apparatus was identical to that used in Exp. I, except that the goal box was a 9 × 15-in. area at right angles to the alley. (Insertion of a sliding panel converted this goal box to the narrow one used in Exp. I.) Goal box walls were 11 in. high. The food cup was the same one employed in Exp. I, located in the same position straight ahead of the 2-in. alley.

Procedure.—The drive maintenance, pretraining, and experimental training procedures were identical to those employed in Exp. I. A more detailed description of methodology and apparatus may be found elsewhere (Goodrich, 1958).

Results[3]

Locomotor performance.—Starting, running, and goal box speeds are plotted in Fig. 2. As in Exp. I, we find for the starting and running measures that the 50% *Ss* ran slower than the 100% *Ss* in the early trials, and faster than the 100% *Ss* in the later trials.

[3] For the purposes of the present paper, Exp. II is a replication of Exp. I with a different sample of *Ss* and different apparatus. A discussion of the effects of goal box size variation may be found elsewhere (Goodrich, 1958).

Analyses were computed using the means over pre- and post-crossover trials as criterion measures. The interaction of trials with reinforcement schedule was significant for both the starting and running speeds ($F = 14.22$, $df = 1$ and 28, $P < .001$, and $F = 16.20$, $df = 1$ and 28, $P < .001$, respectively). Analyses were then computed for the pre- and post-crossover data separately. Prior to the crossover the group differences were nonsignificant ($F < 1$ in both cases), but following the crossover the differences were statistically reliable for both starting and running speeds ($F = 8.68$, $df = 1$ and 28, $P < .01$, and $F = 5.47$, $df = 1$ and 28, $P < .05$, respectively).

For the goal box measures, a very slight early-trial decrement occurred in the curve of the 50% group, but beyond Trial 24 the curves remained together, and no statistical evaluation was attempted. These goal box speed results are not consistent with those of Exp. I. However, a kind of gradient effect was still present in that the initial 50% decrement lasted for greater numbers of trials as performance was measured closer to the food cup. Examination of a plot (Goodrich, 1958) of the 50% data for rewarded and nonrewarded trials separately again precluded any possibility of the relatively slow performance of the 50% *Ss* in the goal box region being due to *Ss* slowing down as they approached the food cup on nonrewarded trials.

Frequency of stops and time spent at the food cup.—The results for these measures are essentially the same as in Exp. I. In each of the three alley segments the 50% *Ss* made more stops than the 100% *Ss* during those early trials on which the corresponding speed data showed the 50% *Ss* to have run slower. Later in training there were no differences between groups, except for a slight tendency for the 100% *Ss* to make more stop responses in the goal box region. Plots of the data (Goodrich, 1958) for time spent at the food cup showed that the 50% *Ss* stayed consistently longer at the food cup throughout training ($t = 1.67$, $df = 28$, for Trials 45–60).

Discussion

The two experiments reported here provide a consistent picture of the effects of 50% vs. 100% reinforcement schedules upon the acquisition of an instrumental response. They indicate that, (*a*) following an early-trial advantage for *Ss* in the 100% group, there occurs a reversal in the case of starting and running speeds so that 50% *Ss* are responding faster at the performance asymptote, and (*b*) the early-trial 50% reinforcement decrement persists for greater numbers of trials as performance is measured closer to the goal region of the runway.

The finding of greatest importance for existing theoretical formulations of instrumental conditioning is that 50%-reinforced *Ss* exhibit a higher level of performance than 100%-reinforced *Ss* at the performance asymptote. An extension of the theoretical schema of Amsel and his students (Amsel, 1958; Amsel & Hancock, 1957; Amsel & Roussel, 1952; Wilson, Weiss, & Amsel, 1955) provides, to a first approximation at least, an adequate interpretation for this phenomenon. According to this formulation, the occurrence of nonreinforcements after some minimal number of reinforced trials produces a motivational state of "frustration" (*F*) which may act to facilitate subsequent performance (Amsel & Hancock, 1957; Amsel & Roussel, 1952; Wagner, 1958). Amsel further postulates that a conditioned (antedating) form of *F*, called $r_f - s_f$, will occur back in the alley itself. If motivational (facilitating) properties are attributed to $r_f - s_f$ (as the formula-

tion seems to call for but Amsel himself does not discuss), the eventual superiority of the 50% *S*s can be interpreted as being due to a motivational increase produced by the occurrence of $r_f - s_f$.

This frustration interpretation is not without its problems. Although the motivational effects of primary frustration (*F*) were clearly demonstrated in studies by Amsel and Hancock (1957), Amsel and Roussel (1952), and Wagner (1958), these facilitating effects did not show up until some 20 to 30 trials had been run. In the present study the advantage in favor of the 50% group began to develop, in the case of starting speeds, as early as the third block of trials (Trials 9–12). It is therefore necessary to hold that conditioned frustration, $r_f - s_f$, had an earlier effect in the present study than did primary frustration, *F*, in the Amsel and Wagner studies. It is quite possible that the narrow width and visual homogeneity of the starting and running sections of the runway used here discouraged competing responses, and thus enabled motivational increases to affect the instrumental running response directly. Greater numbers of competing responses in the early trials could have masked such effects in the Amsel and Wagner studies. A more accurate index of competing responses than the crude one employed here would seem desirable for future study of these problems.

Amsel's frustration formulation attributes inferior performance under 50% reinforcement to the occurrence of competing responses elicited by the stimulus component of $r_f - s_f$. Application of this schema to the present results would require the occurrence of $r_f - s_f$ during the very early trials. However, Amsel postulates that frustration does not occur until previous rewarded trials have resulted in the growth of some minimum amount of incentive motivation, or $r_g - s_g$. Unless it is assumed that the pretraining procedures in the present study were sufficient for the growth of this $r_g - s_g$, it is doubtful whether r_f, and thus s_f, would occur early enough in training to account for the early-trial 50% decrement. An alternative, and perhaps more reasonable, interpretation would be that competing responses occur and hence are learned in the empty goal box on nonreinforced trials, and subsequently interfere with the running response (Adelman & Maatsch, 1955; Weinstock, 1954). The relationship found in the present study between the 50% decrement and the occurrence of stop responses, although only suggestive because of the statistical unreliability of the stop data, is consistent with such a view.

According to Weinstock's (1954) interpretation, which is based on a statistical-contiguity type learning theory plus an ad hoc assumption about the habituation of competing responses, a partial reinforcement schedule results most importantly in the conditioning of competing responses. Weinstock is led to conclude that, although the decrement due to these interfering responses may be small during the later acquisition trials (because of habituation), final asymptotic levels of running speed should be in direct relation to the percentage of reinforcement employed. Although competing responses may account for an early-trial 50% decrement, Weinstock's formulation is clearly inadequate as it stands to account for the fact, indicated by the available evidence, that partially reinforced *S*s exhibit a higher final performance level than continuously reinforced *S*s.

In both of the experiments reported here, partial reinforcement effects were found to vary with the portion of the response chain being measured, a finding consistent with the Freides (1957) and Haggard (1956) studies. No clear interpretation of the effect seems available at present. Similarly, the interpretation of the apparent relationship between the time spent at the food cup and speed of running must await additional research. The present results are ambiguous in that shorter time may represent either a faster rate of eating or merely picking up the food and moving away from the cup without actually eating the food. A measure

of actual consummatory rate could be obtained more easily with a liquid incentive.

Summary

Two experiments were made to demonstrate that the relative performance of 50% and 100% reinforcement groups varies during the course of acquisition, and that these effects vary with the point in the response chain measured. In each experiment, two groups of 15 female hooded rats were given 60 acquisition trials in a straight alley, one group receiving 50% partial reinforcement and the other 100% reinforcement. The experiments differed only in the size of the goal box used.

The results of the two experiments agreed in showing that for the starting and running speed measures the 50% group was inferior in the early trials but superior at the performance asymptote. For the goal box speed measure no 50% superiority developed; the 50% *Ss* showed slower performance throughout training in Exp. I, and did not differ from the 100% *Ss* in Exp. II. In both experiments the 50% decrement lasted for greater numbers of trials as performance was measured closer to the goal region of the apparatus.

The results were discussed in relation to Amsel's frustration theory, although it was concluded that no existing theoretical schema provides an entirely adequate interpretation.

References

Adelman, H. M., & Maatsch, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Amsel, A., & Hancock, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

Finger, F. W. The effect of varying conditions of reinforcement upon a simple running response. *J. exp. Psychol.*, 1942, **30**, 53–68. (a)

Finger, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **31**, 120–133. (b)

Freides, D. Goal box cues and pattern of reinforcement. *J. exp. Psychol.*, 1957, **53**, 361–371.

Goodrich, K. P. Acquisition of an instrumental response as a function of reinforcement schedule and goal box confinement. Unpublished master's thesis, State Univer. of Iowa, 1958.

Haggard, D. F. Acquisition and extinction of a simple running response as a function of partial and continuous schedules of reinforcement. Unpublished doctoral thesis, State Univer. of Iowa, 1956.

Jenkins, W. O., & Stanley, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

Lewis, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *J. exp. Psychol.*, 1956, **51**, 45–53.

Lewis, D. J., & Cotton, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, **50**, 189–194.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

Wagner, A. R. Motivational effects of nonreinforcement as a function of the reinforcement schedule. Unpublished master's thesis, State Univer. of Iowa, 1958.

Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

Wilson, W., Weiss, I. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–61.

(Received March 16, 1958)

Journal of Experimental Psychology
Vol. 57, No. 1, 1959

REPLICATION REPORT: MEANING ESTABLISHED BY CLASSICAL CONDITIONING[1]

ARTHUR W. STAATS, CAROLYN K. STAATS, WILLIAM G. HEARD, AND LARRY P. NIMS

Arizona State College, Tempe

It was shown in a previous study (Staats & Staats, 1957) that word meaning could be conditioned to nonsense syllables. The long procedure was monotonous and *S*s were run in a group. The present study attempts to replicate the original study using individually run *S*s in a much shortened procedure. In addition, multiple *E*s were employed in the present study.

Method.—The general procedure has been described in the prior study. In contrast to the earlier study, four nonsense syllables were used (instead of six): YOF, XEH, JIC, and QUG; and the two preliminary learning tasks were deleted. Two experimental conditions were employed in a counterbalanced design. Half the *S*s (Group I) were positively conditioned to XEH and negatively to YOF. This was reversed for the other *S*s (Group II). QUG was paired, for all *S*s, with a class of words denoting building materials and JIC with "transportation" words. This was different from the previous study in which the filler syllables were paired with words of unsystematic meaning. Each syllable was paired with 14 different words, i.e., there were 14 conditioning trials instead of the 18 in the original study. The *S*s were told that the purpose of the experiment was to study two different types of learning and that their task was to separately learn the syllables and words which were presented simultaneously. Meaning of the syllables was measured as in the previous study. Finally *S*s listed all the words they could remember, and were questioned regarding the purpose of the experiment.

TABLE 1

MEANS AND *SD*'S OF CONDITIONED MEANING SCORES

Group	Syllables			
	XEH		YOF	
	Mean	*SD*	Mean	*SD*
I	2.92	1.75	4.33	2.21
II	5.67	1.65	1.97	0.86

Note.—The pleasant pole scored 1, unpleasant 7.

Subjects.—The 24 new *S*s were randomly drawn from the same population as in the first study and were randomly assigned to one of three *E*s (8 *S*s per *E*), and to one of the experimental conditions.

Results.—The score for the extent of conditioning of *S* was obtained by subtracting the semantic differential score of the positively conditioned syllable from that of the negatively conditioned syllable. The mean conditioning score for the undergraduate student *E* was 2.25, for the graduate student 2.5, and for the faculty member 3.00 (simple analysis of variance $F < 1$). The remainder of the analysis was the same as in the previous experiment. The means and *SD*'s are presented in Table 1. An analysis of variance revealed that conditioned meaning responses were significant ($F = 18.62$, $df = 1$ and 22, $P < .001$). The differences between nonsense syllables ($F = 3.80$, $df = 1$ and 22) and between groups ($F < 1$) were not significant. Some *S*s realized certain syllables were paired with certain types of words, but none was aware of the relation between this fact and scoring of the semantic differential scale.

Discussion.—Strong language conditioning of meaning was obtained with relatively few *S*s, using three *E*s differing in age and status, when *S*s were individually run in a shortened procedure (time of procedure was cut from 45 min. to 20 min.). The differences between the results of the different *E*s were not significant. The present procedure should be equally effective using individual *S*s or groups of *S*s.

REFERENCE

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

(Received May 29, 1958)

[1] This study is part of a project sponsored by the Office of Naval Research under Contract Nonr 2305 (00) with Arizona State College at Tempe.

BIOLOGICAL AND BIOCHEMICAL BASES OF BEHAVIOR

Edited by Harry F. Harlow and Clinton N. Woolsey

Contained in this volume are the papers presented at the Symposium on Interdisciplinary Research, held at the University of Wisconsin. The purpose of the symposium was to correlate studies in progress in the fields of anatomy, physiology, biochemistry, and behavior involving the research of numerous laboratories. This collection represents a significant step forward in the growth of interdisciplinary research.

496 pages $8.00

The University of Wisconsin Press

430 STERLING COURT MADISON 6, WISCONSIN

A *very considerable* proportion of all published articles on perceptual-motor skills and on perception-as-influenced-by-experience are given immediate publication in

PERCEPTUAL AND MOTOR SKILLS

If you are interested in these fields, you and your library should have current subscriptions to PMS ($10.00 per year, with a special rate of $5.00 to individual members of the APA). Prices of back volumes may be obtained on request from

PERCEPTUAL AND MOTOR SKILLS
PSYCHOLOGICAL REPORTS
Box 1441
Missoula, Montana